Investigating the Impact of Urbanisation on Rainfall-Runoff Models

Fidal, James

Award date:
2019

Awarding institution:
University of Bath

Investigating the Impact of Urbanisation on Rainfall-Runoff Models

James Fidal

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Architecture and Civil Engineering

March 2019

Acknowledgements

Firstly, I would like to express my gratitude to my supervisor Dr. Thomas Kjeldsen for the continuous support of my Ph.D study and related research, knowledge and his patience. His insightful comments and criticism helped at all times during my research and writing of my thesis.

I would like to thank the Engineering and Physical Sciences Research Council for my funding grant. Alongside the NERC funded POLLCURB project for providing access to hydrological and land-use data. I would also like to thank Hyeonjun Kim and Cheolhee Jang of KICT for providing access to the CAT model, the hospitality when visiting South Korea and the interesting discussions. Of the University of Bath I would like to thank everyone apart of the Research Unit for Water, Environment and Infrastructure Resilience for providing many insightful discussions.

Last but not least my parents and sister for supporting through my research and writing of my thesis. Of my friends Sam, Luke, Mark, Joe and Aled. Also Huw Thomas and Will Rowland for providing a single comment on my thesis I thank them.

Abstract

The introduction of urban land-cover changes the hydrological characteristics of catchments. So there is a need to develop new methods to account for this change in land-use. Whilst a large number of hydrological models exist, each are designed for different purposes. However numerous sources of literature highlights the shortcomings of urban rainfall-runoff models. The aim of this thesis is to create a generic framework for modelling the impact of urbanisation on catchment runoff using lumped conceptual rainfall-runoff models.

A number of urban runoff generation frameworks are designed and tested within this thesis. The best performing framework uses a newly developed scaling term which can account for the variability of infiltration in urban areas, when combined with an urban routing framework produces the URMOD model.

The URMOD model is extensively tested on multiple urban catchments using both daily and sub-daily data. The model performance was gauged using the model comparison tools developed within this thesis, a jackknife calibration/ validation method, paired Z-test and a binomial hypothesis test. The comparisons took place on a number of catchments in the Thames catchment and the Gyeongan-Cheon catchment. Whilst some performance issues are raised for URMOD, results show that URMOD was able to better represent the urban effects on catchments than a rural model and an existing semi-distributed model when using daily and sub-daily data.

Contents

List of Figures

List of Tables

Chapter 1

Preface

The Royal Meteorological Society's president Sir Bernard Augustus Keen gave an address in 1939 entitled "What happens to the rain?" Pereira (1982). Despite this being a simple question, the answer is incredibly complex due to the dynamic nature of hydrology. Whilst hydrological knowledge has advanced since 1939, both in practice and theory, this question is still relevant today.

Estimating stream flow has been attempted for over 100 years, dating back to 1850 with Mulvany, who used a rational method to attempt to estimate storm runoff (Mulvaney, 1951). In the 1930s the concept of using a simple unit hydro graph to investigate runoff was presented (Sherman, 1932). From then numerous different methods were developed in an attempt to quantify the relationship between rainfall and runoff, and to predict future runoff. One method to capture the effects of rainfall on stream flow is rainfall-runoff modelling. A rainfall-runoff model is a mathematical model which attempts to describe the relationship between rainfall and runoff within a specified region. More specifically Devia et al. (2015) defines a rainfall-runoff model as a set of equations that help estimate runoff as a function of various parameters used for describing catchment characteristics. These regions are primarily catchments, defined as the area of land in which water drains from before flowing into a water system, e.g. rivers and lakes. The general components that make up a mathematical model are a set of model processes or governing equations, coupled with a set of inputs which generates a set of output data (Singh et al., 1995). Typical inputs of rainfall-runoff models are hydrological data such as observed rainfall, observed river runoff and climate data such as evaporation, with typical outputs simulated river flow for a catchment outlet. As outlined in Salvadore et al. (2015) and Beven (2011), there are various types of rainfall-runoff models, each designed for a different purpose, such as developing or testing hydrological theories, attempting to understand the hydrology of a catchment or prediction for water management. As outlined by Devia et al. (2015), there is a wide range of rainfall-runoff models, from very detailed distributed models that can capture the spatial distribution of inputs to model the flow throughout the catchment to lumped models attempting to generalise the entire catchment as a single entity. Thus, care has to be taken in order to select the correct model for the task at hand.

Some of the most significant impacts on the environment are due to humans, and one such influence is urbanisation since currently more than 50% of the world's population are living in urban areas (WHO, 2014), with projections showing an increase to 60% by 2030. Urbanisation typically leads to an increase of impervious surfaces as previously pervious surfaces such as soil and grass are paved over to make streets and houses. This drastic change in environment has significant consequences on the hydrology of the area. As outlined by Niemczynowicz (1999), Fletcher et al. (2013), Miller et al. (2014), DeFries and Eshleman (2004), Rosso and Rulli (2002), Jones et al. (2000) and Shuster et al. (2005) urbanisation and the process of sealing previously pervious surfaces leads to a reduction of infiltration of the pre-existing soil and an increase in surface runoff. Not only is there an increase in surface runoff but transportation of runoff to the rivers is much faster due to pipe networks and precipitation travelling faster over impervious surfaces. All of these factors brings a new layer of complexity to the initial question “what happens to the rain?”. A method to capture these changes is rainfall-runoff modelling, but with models that can account for urbanisation by simulating urban processes such as pipe networks or transpiration over sealed surfaces. There are a wide range of models considering urban impacts on the hydrological cycle, ranging from detailed hydrologic models that attempt to simulate flow in every single pipe in a network, up to conceptual catchment models that try to generalise the hydrology of the entire impervious areas. One problem with every rainfall-runoff model is that they take generalisations of catchment processes, such as lumping topographical features together. As such they are only an approximate representation of the true hydrological processes of a catchment. This problem will be explored in more detail in Section 1.1.

1.1 Problems with Models

The observed values of river flow (q_{obs}) differs from the notationally true value (μ) by some observational error (ϵ_1) (Beven, 2011), as expressed in Equation 1.1:

$$q_{obs} = \mu + \epsilon_1. \tag{1.1}$$

Hydrological models can be used in an attempt to simulate catchment runoff q_{sim}. Consider a mathematical model M with input data I and estimated parameters $\hat{\theta}$ which simulates catchment runoff as:

$$q_{sim} = M(\hat{\theta}|I). \tag{1.2}$$

Since $\hat{\theta}$ is estimated through calibration of the models based on an observed sample of finite duration there will be some error due to differences between the optimum parameters θ and the estimated parameters $\hat{\theta}$. There will be circumstances when θ is not a unique set, due to optimum parameters not being able to be obtained because M is an approximate representation of the hydrological processes within the catchment. Since M is a representation of the catchment, when the optimum parameters (θ) are known the simulated runoff will still have model structure error ϵ_2, as shown in Equation 1.3:

$$\mu = M(\theta|I) + \epsilon_2. \tag{1.3}$$

By combining Equation 1.1 and Equation 1.3 the observed runoff can be expressed as Equation 1.4:

$$q_{obs} = \mu + \epsilon_1 = M(\theta|I) + \epsilon_1 + \epsilon_2. \tag{1.4}$$

The observed riverflow is therefore a sum of the model plus two error terms, the model error ϵ_2 and the initial observation error ϵ_1. Taking the residuals as the difference between the observed and simulated, Equation 1.5 is obtained:

$$(q_{obs} - q_{sim}) = M(\theta|I) + \epsilon_1 + \epsilon_2 - M(\hat{\theta}|I) \tag{1.5}$$
$$= M(\theta) - M(\hat{\theta}) + \epsilon_1 + \epsilon_2 \tag{1.6}$$
$$= \epsilon_1 + \epsilon_2 + \epsilon_3. \tag{1.7}$$

The problem looking at this result is that a comparison of the residuals will include uncontrollable observed ϵ_1 and model ϵ_2 errors. Alongside these two errors a third error is present, due to the most optimal set of parameters not being known ϵ_3,

$$\epsilon_3 = M(\theta) - M(\hat{\theta}). \tag{1.8}$$

Whilst not addressed in this thesis, input I can be uncertain i.e. the true value is not known. When this is the case the previous formulation will still be true except a new error ϵ_4 will have to be introduced to account for the input uncertainty. ϵ_1 will not be considered in this thesis, due to this thesis being concerned with reducing model error not reducing data observational error. Data quality will be checked but observation error ϵ_1 will be ignored and the assumption that the data is the best quality it can be will be taken. As such the two errors considered in this thesis are ϵ_2 and ϵ_3. Details of how these errors are considered are highlighted in Section 1.2.

1.2 Research questions

Following from Section 1.1 the overall aim of the research is to develop a more accurate rainfall-runoff model structure M that better represents urban effects, which will reduce the model error ϵ_2. The model created will be a conceptual model which try to take a generalised approximate representation of the true hydrological system. The model designed and created is called URMOD. To answer the overall aim of the research a number of research questions are formed and answered throughout the thesis.

Research question 1: Can rainfall-runoff modelling in urban catchments be improved via implementation of a conceptual urban framework onto an already existing rural based model ?

The first research question will attempt to reduce model error ϵ_2 via developing a simple, but generic, urban framework extension in order to better represent the urban processes within urban catchments. This will be explored in Chapter 3.

Research question 2: What is the loss of information when moving from lumped to spatially distributed rainfall-runoff models on a varying time scale for urban catchments ?

The second research question will explore model performance between two different models, in order to explore the differences for each ϵ_2 created by the models. The focus will be exploring how changing time scales and level of spatial detail impacts model performance and subsequent model error ϵ_2.

Research question 3: Can simple statistical-based techniques be used to improve validation of rainfall-runoff models.

The third research question attempts to fill the gap that either very basic or more advanced techniques are used to validate rainfall runoff models. The third research question will be achieved via the development of a more robust model calibration and validation which will attempt to reduce the error obtained from model parameters not being the optimal set ϵ_3.

1.3 Thesis structure

The thesis consists of seven chapters. Chapters 4 to 6 are presented in the form of papers. The list below summarises the content in each chapter.

Chapter 1 is an introduction to the thesis, presenting the general context and themes addressed throughout the thesis. With an outline of the overall aim and research questions of the thesis.

Chapter 2 is a literature review, considering the main approaches to urban rainfall-runoff modelling, highlighting the gaps in knowledge. The short comings on model performance techniques will be reviewed.

Chapter 3 describes in detail the urban model structure designed for this thesis (URMOD). This model structure is the foundation of the subsequent model development chapters, alongside the model that will be used to answer research questions 1 and 2.

Chapter 4 describes the model calibration method and performance techniques designed for this thesis. With a case study showcasing these techniques, through a comparison of an urban model presented in Chapter 3 and a rural rainfall-runoff model. This chapter will focus on answering research question 3.

Chapter 5 presents a comparison between an urban rainfall-runoff model URMOD and a rural rainfall-runoff model described in Chapter 3. Utilising the techniques developed in Chapter 4 the models are applied to two urban catchments in the River Thames Basin using sub-daily data. This chapter will focus on research question 1.

Chapter 6 presents a comparison of two urban models, the first URMOD described in Chapter 3, the second the existing Catchment Assessment Tool (CAT model). Using the techniques developed in Chapter 4, the models are applied to two urban catchments. The first located in South Korea, the second located within England within the River Thames Basin. Both comparisons are taken place using sub-daily data. This chapter will focus on research question 2.

Chapter 7 is a conclusion to the thesis, with a discussion and conclusions highlighting the answers to the research questions and aim presented in Chapter 1.

Chapter 2

Literature Review

2.1 Part 1: Background and introduction to hydrological modelling

2.1.1 Human impact on catchment hydrology

As of 2015 54% of the total global population was living in urban areas (WHO, 2014), with projections showing an increase to 60% by 2030. The resultant increases to impervious surfaces leads to a reduction in infiltration which can have considerable impacts on the hydrology of catchments, (Niemczynowicz, 1999), typically leading to a higher percentage of precipitation becoming surface runoff (Shaw, 1994). One important problem with increasing urban catchment runoff is this can lead larger flood returns (Houston et al., 2011).

Alongside a reduction in infiltration into the soil, urbanisation brings changes to river flow. As reported by Ahn and Merwade (2014), who use a collection of methods such as trend analysis, hydrological modelling and impact quantification in order to conclude that human activates have a larger impact on increasing flow change than that of climate change. Yang et al. (2010) suggested that impervious surface area as low as 3%-5% in catchments is the threshold, over which urbanisation starts to have a detectable influence on river flow. The review by Fletcher et al. (2013) on urban hydrology highlighted that increasing urbanisation leads to higher runoff rates in catchments. This is supported by Jones et al. (2000) who reported the impacts of the increasing number of roads on peak flows of streams, concluding roads lead to an increase in flood peaks. Similarly Miller et al. (2014) concluded that increases in impervious land cover for rural catchments can have a much greater impact on peak flows for catchments. These conclusions are shared by Rose and Peters (2001) who found that peak flows were increased due to urbanisation in Atlanta USA. Hawley and Bledsoe (2011) investigated urbanisation in southern California, and found that catchments with 20% impervious land cover experience five times as many days of mean daily flows on the order of 100 m^3/s and three times as many days on the order of 1000 m^3/s relative to rural catchments.

Alongside this increase in river flow it has also been shown that the urban area of a catchments runoff leads to the river quicker than that of rural catchments. For example Shaw (1994) concluded that a consequence of increasing urbanisation in a catchment is that the lag time between precipitation and river runoff is decreased. This is further backed up by Shuster et al. (2005) who stated that increases in impervious surfaces can result in a shorter lag time between initial rainfall and runoff resulting in higher runoff peaks. In addition to more runoff and higher flows being generated, urbanisation can bring changes to other parts of the stream dynamics. Low flows in urban catchments are decreased due to reduced contributions from groundwater storage; due to reduced infiltration. Brun and Band (2000) explored the impact of increasing urban land cover in upper Gwynns Falls in USA found a 20% decline in base flow in a 17-year period between 1973 and 1990. This decline in base flow is further explored in Bhaskar et al. (2015), who concluded that impervious surface cover only led to a 0.56% increase in subsurface storage which resulted in decreased groundwater discharge as base flow. Rosso and Rulli (2002) concluded that land-use change did not have an important impact on the Bisagno river, reporting only a negligible impact on the upper section of the river, and a moderate impact on the lower river, but a more severe impact on the tributaries that drains the urban areas.

Diem et al. (2018) highlighted the diverse impacts that urbanisation can have on stream flow. The study explored eight catchments in Atlanta in which all catchments had experienced increased urban land cover over the past 30 years. They concluded that urbanisation had different impacts on the catchment. For six catchments high-flow events were increased when no increase in annual rainfall was recorded for the same period of time. Large increases in stream flow occurred in two watersheds, which transitioned from forested land to developed urban land. In a few catchments stream flow actually declined despite urbanisation.

From the literature, it is evident that urbanisation is generally found to have an impact on stream flow, by reducing infiltration rates, which lead to more runoff or simply changing the stream by a reduction in base flow. Alongside the reduction in baseflow a decrease in lag-time between rainfall and runoff events for urban areas has been observed in a number of studies. All of these changes have important impacts to not only the river hydrology but the surrounding catchments, for example increasing runoff can lead to larger occurrence and volume of floods which can lead to large repair costs. Whilst this will not be addressed in the thesis, chemical and quality impacts are also associated with urbanisation as runoff can transport harmful materials into river systems. Methods are needed in order to attempt to explain the changes within the catchments due to urbanisation.

2.2 Part 2: Review of models

2.2.1 Introduction to hydrological modelling

A rainfall-runoff model is a mathematical model which attempts to describe the relationship between rainfall and runoff for a specified region. Rainfall-runoff modelling plays a key role in water management, as these models can be used to expand stream flow records, predict future flows and predict flows in ungauged catchments (Beven, 2011). Hydrological models vary in complexity, each with a different purpose and answering different hydrological questions e.g. Shoemaker (1997), Salvadore et al. (2015). Consequently models can be divided into different types as will be further explored and reviewed in Section 2.2.

As discussed by Džubáková (2010), Beck (1991) and Beven (2011), the majority of rainfall-runoff models can be split into three types with alternative names in brackets: (i) metric (data-based, empirical or black box), (ii) conceptual (parametric, soil moisture accounting or grey box) and (iii) physical-based (white box or mechanistic). Alongside these three types, models can be split again based on how much spatial data they can accommodate, these classifications being: (i) lumped, (ii) semi-distributed, and (iii) distributed models. Section 2.2.2 will provide definitions to the different types of rainfall-runoff models focusing on advantages and disadvantages. Section 2.2.3 will provide definitions for the different types of spatial models, again providing advantages and disadvantages of each type. Finally Section 2.2.4 will present a comparison between the types of models, discussing existing literature on the comparison of spatially varying models.

2.2.2 Classification of rainfall-runoff models

The complexity of a rainfall-runoff's model structure determines how runoff can be calculated. This section will briefly describe the three main types of model structures; Metric, conceptual and Physically-based. Each description will outline the data requirements and present some example models.

Metric models

Wagener et al. (2004) define metric models as those that are observation based and that create system response by utilising information from existing data. These type of models are primarily designed with little or no consideration of hydrological system processes. Metric models typically do not have any spatially varying hydrological processes, thus treating the catchment as a single lumped entity. One advantage of the Metric model is they do not require many different types of data, but do require access to long and reliable runoff data. However one drawback of Metric models as outlined by M and J (2004) is that because these models rely on certain climate conditions under which the data is collected, these models cannot easily be generalised for use on other catchments. One of the earliest form of Metric models is the unit hydrograph theory (Sherman, 1932); a linear model which can be used to obtain the hydrograph from excess rainfall in a catchment. More recently artificial neural network models have been used, originally presented by McCulloch and Pitts (1943). Artificial neural networks have been applied for problems such as: river flow prediction (Wright et al., 2002), (Riad et al., 2004) and (Smith and Eli, 1995) and prediction of evaporation (Sudheer et al., 2002).

Conceptual models

Conceptual (also named as parametric) models are designed prior to use to attempt to represent hydrological processes through simplified conceptualisations of hydrological processes Beven (2011). For example, by using a number of connected reservoirs that represent physical characteristics of a catchment such as the semi-distributed parallel surface rainfall-runoff conceptual model (Hsieh and Wang, 1999). These reservoirs are filled through rainfall, infiltration and emptied via evaporation, drainage etc. As such, more observed catchment data is needed than for metric models such as catchment area or land-use data but, due to the existence of a model structure, conceptual models can more easily be generalised for multiple catchments. Conceptual models require meteorological and hydrological data for model parameter calibration. However, one issue with this type of modelling is that an increase in model processes can lead to an increasing number of model parameters which can lead to overparameterisation (Perrin et al., 2001). One of the first conceptual models was the Stanford watershed model (SWM) developed by Crawford and Linsley (1966). More recently developments include models such as MODHYDROLOG (Chiew and McMahon, 1994), Hydrologiska Byråns Vattenbalansavdelning (HBV) model (Bergstrom, 1976) and TOPMODEL (Beven et al., 1984).

Physically-based model

Physically-based models, attempt to represent hydrological processes through the understanding of the physical laws governing the hydrology processes (Pechlivanidis et al., 2011). Governing equations derived from first principles are applied in order to represent multiple parts of the hydrological cycle. These equations typically include conservation of mass and energy, kinematics and water balance equations. The parameters for physically-based models do not normally require calibration but instead represent the physical characteristics of a catchment, for example soil properties, elevation and dimensions of a river. Hence, these models do not ideally require more hydrological/meteorological data than conceptual models but require physical measurements from catchments, which can lead to extra data demand, scale issues and over-parameterisation. Models such as the MIKE SHE (Abbott et al., 1986), Water and Energy Transfer between Soil, Plants and Atmosphere (WetSpa) (Wang et al., 1996) and United the Soil and Water Assessment Tool (SWAT) (Arnold et al., 1998) are considered physically based.

2.2.3 Classification of spatial representation by models

Alongside the three classifications of models presented in Section 2.2.2, models can represent spatially varying hydrological processes in different ways. Thus creating three new classifications, lumped, semi-distributed and fully distributed models.

Lumped models

A lumped model assumes the entire catchment as a singular unit and as such takes no spatial variation of the hydrological processes into account. These types of models use averaged values for parameters across the entire catchment, and consequently only a singular river flow value at the catchment outlet is calculated. As lumped models require averaged values limited amounts of input data are needed, but this is at the expense of model resolution as detailed spatial variations cannot be represented by this type of models. One important limitation of this assumption is that lumped models cannot easily represent large catchments accurately, especially if the catchment is diverse with elevation changes and multiple different soil types as outlined by Moradkhani and Sorooshian (2009) and Uhlenbrook et al. (2004). However, one advantage of lumped models is that computation time can be less due to hydrological processes from a singular area being considered. Previous published lumped models include the Identification of unit Hydrographs And Component flows from Rainfall, Evaporation and Streamflow (IHACRES) model (Jakeman and Hornberger, 1993), Australian Water Balance Model (AWBM) (Boughton and Chiew, 2007) and Génie Rural à 4 paramètres Journalier (GR4J) (Perrin et al., 2003).

Semi-distributed models

Semi-distributed models as defined by Beven (2011) are the set of model that do not make simulations throughout the entire catchment but rather for a distribution function of characteristics. Therefore semi-distributed models tend to have more parameters than lumped models as each sub-catchment has its own set of parameters. A clear advantage of semi-distributed models is that the sub-catchments can represent spatial variation of: (i) climate, (ii) soil types, (iii) elevations or (iv) land-use. However one potential disadvantage is that semi-distributed models require access to more data than lumped models. Examples of semi-distributed models include TOPMODEL (Kirkby and Beven, 1979), The Soil and Water Assessment Tool (SWAT) (Arnold et al., 1998), RAFTS (Goyen et al., 1991) and the Catchment Assessment Tool (CAT) (Kim et al., 2012).

Fully distributed models

Fully distributed models, split the catchment into a grid of smaller cells which can be square cells (rasters), triangular irregular networks (TIN) or irregular shaped objects and then model the flow of water through each cell. Examples of application where the use of distributed models is considered useful are highlighted by Kampf and Burges (2007) and include: (i) pollutant transport, (ii) hydrologic responses to land-use changes and (iii) sediment transport. One advantage of this type of model is that flow can be modelled at the outflow of each grid cell throughout the entire catchment as opposed to the other methods of modelling which take snapshots at certain outlets. However, one major disadvantage of these models is the extensive data requirements in order to model every grid cell. However if data is available this type of model can provide the most detailed hydrological description of catchments. Available distributed models include the MIKE SHE (Abbott et al., 1986), Water and Energy Transfer between Soil, Plants and Atmosphere (WetSpa) (Wang et al., 1996), distributed object-based rainfall-runoff simulation model (DORS) (Zhou et al., 2010) and geomorphology-based hydrological model (GB) (Yang et al., 2000).

2.2.4 Comparison of spatial representation models

The choice between a lumped, semi-distributed or fully distributed model is prevalent thought-out the hydrological literature, with conflicting conclusions. Comparing the different model structures is important in order to determine optimal model structure from the large array of existing models (Clark et al., 2011). As stated by Smith et al. (2004) whilst comparisons of hydrological models on river flow have been conducted, no comprehensive comparison of lumped and distributed modelling methodology had been published by 2004. As such the distributed model intercomparison project (DMIP) (Smith et al., 2004) was designed primarily to offer guidance to the US National Weather Service on future modelling strategies alongside attempting to explore the spatial variability of rainfall and its effects on catchments. The DIMP project was a comparison of a large number of distributed models and a single lumped model in order compare multiple model frameworks in order to determine which model structure performed the best. The results of the DIMP project were presented in Reed et al. (2004), showing that the lumped model outperformed the distributed model in more instances than the reverse. However the distributed models performed better on larger catchments than the lumped model.

Summary

The main conclusion that can be drawn from Section 2.2 is that selecting an optimal model structure depends on the specific hydrological question that needs to be answered and the amount of data available. For example, if flow within a catchment needs to be explored then a lumped model should not be chosen. Results from Table 2.1 and results from the DIMP project indicate that if detailed spatially varying data is available then distributed models are preferred to lumped models. But when a singular lumped data set is available then semi-distributed or lumped models can match and even perform better than distributed models. One issue apparent from the literature is that it is not generally possible to fully parameterise distributed and physically based models due to the large data requirements, resulting in the need for accurate lumped models for catchments when the aforementioned models cannot be fully parameterised.

2.3 Part 3: Modelling urban hydrological processes

As outlined in Section 2.1.1 sealing of impervious areas has considerable impact on the hydrology of an area. A method to capture these effects is a rainfall-runoff model which can account for urban processes. However, there are conflicting conclusions in the literature on how best to model runoff generation in urban areas

Models	Reference	Catchment	Result
MIKE SHE, TOPMODEL and geomorphology-based hydrological model GB model (All Distributed)	Yang et al. (2000)	Seki River catchment in Hokuriku region of Japan (703 km^2)	Performance was similar but each model had advantages. The MIKE SHE was preferable for comprehensive studies on smaller catchments due to its ability to represent spatial information. The TOPMODEL was preferable for rainfall runoff-simulations, and the GB model for complete simulations of larger catchments.
Three versions of HBV model structure: (one lumped, two semi-distributed, one distributed)	Das et al. (2008)	Upper Neckar catchment in southwest Germany. (13 sub catchments ranging from $119.89 km^2$ to $454.65 km^2$.)	Both semi-distributed models outperformed the others.
Three versions of HBV model structure: (one lumped, one semi-distributed, one distributed)	Krysanova et al. (1999)	Eight urban sub-basins of the Elbe catchment as well as the entire 80 657 km^2 Elbe catchment in Germany	Model performance was good using all structures, the semi-distributed and distributed model did perform better when more detailed land cover characteristics were applied.
NAM (lumped), WATBAL (Semi-distributed), MIKE SHE (distributed)	Refsgaard and Knudsen (1996)	Three catchments in Zimbabwe using average daily flow data (254 km^2, 1040 km^2, 1090 km^2)	Similar performance of models. Lumped model performance is variable depending on amount of calibration data. No calibration data distributed models perform better.
SWAT model (Semi-distributed), MIKE SHE (distributed)	El-Nasr et al. (2005)	Jeker catchment in Belgium using daily data (465 km^2)	Fully distributed model was better able to model the overall variation in the river flow.
Five HBV model (lumped, semi-distributed, grid-model without routing, grid-model with hillslope routing and a grid-model with both hillslope and channel routing.)	Li et al. (2015)	Glomma River catchment in Norway (18,932 km^2).	Increasing model complexity did improve model performance. But no differences between the full grid-models was found.
Three Sacramento Soil Moisture Accounting (SAC-SMA) (one lumped, two semi-distributed)	Ajami et al. (2004)	Illinois River basin at Watts catchment (1645 km^2)	At the catchment outlet the model performances were similar.

Table 2.1: Comparison of lumped, semi-distributed and distributed models present in the literature

and how to rout urban runoff to the catchment outlet. Consequently multiple methods have been proposed and this section will review them.

2.3.1 Urban runoff generation methods

Packman (1980) outlined four challenges involved in modelling the effect of urbanisation on flood magnitude, and these challenges also represent generic challenges of modelling urban hydrology: (i) An increase in the complexity of the hydrological cycle. Effects such as the pattern of development, how much change there is due to urbanisation and the type drainage systems all impact the increasingly complex hydrological cycle. (ii) The problems of parametric description of urban development. (iii) A general lack of reliable data from urban catchments. (iv) Difficulty in generalising the results. Such as estimation in ungauged catchments or future conditions in gauged catchments, alongside how best to apply the method across multiple catchments. The focus of this literature review and thesis will primarily address problem: (i), (ii) and (iv). The issue of increased complexity in the hydrological cycle due to urbanisation is still present throughout the literature today over 30 years since Packman (1980) addressed this. A review by Salvadore et al. (2015) concluded that there is still much work needed in order to improve modelling of urban systems, and presented a list of shortcomings such as: the complexity of physical systems leading to a high uncertainty, limited data and inconsistent levels of detail.

All of these factors have led to a large number of hydrological models, proposing different methods for representing hydrological processes in urban areas. Shields and Tague (2012) highlight that currently no universally accepted characterisation of urban surfaces exists for hydrological modelling. Fletcher et al. (2013) and Jacobson (2011) further highlight this claim, stating that most rainfall-runoff models are rarely parametrised to fully account for urban processes other than applying an impervious surface extension to account for an increase in runoff from impervious areas. In the extensive review presented by Salvadore et al. (2015) 30% of the models reviewed accounted for only the impact of impervious cover. This issue is increasingly complex when determining which of the urban processes have the largest impact on the hydrological response, as explored in Section 2.1.1, and raises the issue of how to generalise these urban processes for rainfall-runoff modelling between catchments. Alongside this, currently no single best method exists in order to model the runoff from urban areas as reviewed below.

There are multiple different methods to model the hydrological behaviour of urban surfaces. A common approach is to assume a constant percentage runoff (or infiltration) from urban areas, such as 70% runoff generation, (Packman, 1980) and (Kjeldsen, 2009), or assuming zero infiltration e.g. (Wiles and Sharp, 2008). Rather than using a lumped value for the entire catchment, spatially varying values can be obtained in order to have different runoff values at different parts of the catchment (Valeo and Moin, 2000). Simiraly, Zhou et al. (2010) uses a distributed object-based rainfall-runoff simulation model (DORS) which uses multiple spatial runoff values in order model the urban flow throughout the catchment. Verbeiren et al. (2013) also use spatially-distributed runoff maps, in order to estimate runoff in the Tolka River catchment in east Ireland using the WetSpa model. Another technique outlined by Franczyk and Chang (2009) involves estimating hydrological characteristics of urban areas as a function of proximity to streams.

Whilst there are a number of methods to model urban runoff generation, the fundamental question still remains, how much runoff is generated from impervious surfaces? This question has been highly debated within the literature with conflicting conclusions on the answer. Ragab et al. (2003) stated that lack of accurate data has meant that hydrologists typically assume urban infiltration to be zero and runoff to be 100%. Their study explored runoff values from asphalt roads and concluded the ratio of runoff to rainfall was 70% annually, 90% for winter months (October-March) and 50% for the summer (April-September). The study conducted by Stephenson (1994) indicated that runoff from a suburban catchment in Johannesburg could be as low as 15% of rainfall, with results showing that in the undeveloped areas runoff was as low as 4%. Wiles and Sharp (2008) conducted runoff experiments on asphalt roads which showed that runoff was approximately 79%. Similarly the review by Redfern et al. (2016) presented observational evidence from a range of studies that infiltration through urban surfaces vary depending on the surface type, with values ranging from 16% (deteriorated Asphalt concrete) runoff to 93% (Inclined concrete slab). Redfern et al. (2016) concluded that soil moisture in urban areas can behave differently than rural areas such that the retention of soil moisture is decreased. However, as indicated by Redfern et al. (2016) there is little literature on how best to model urban soil moisture.

This conclusion is echoed by a number of studies, for example Law et al. (2009) highlighted that in modelling literature was limited on characterisation of distributed soils in urban areas. Concluding that the existing research states that the properties of urban soils are significantly different than rural soil properties. Due to the fact that urban soils are compacted more than that of the rural soils which can result in higher runoff values. Similarly Gregory et al. (2006) concluded that soil compaction had a negative impact on infiltration rates in north central Florida. However they highlighted the importance of soil infiltration rates as over or under estimation can lead to under or over estimation in predicted runoff. Lastly, the authors concluded that measuring soil compaction was very time consuming and suggested that if multiple catchment soil infiltration rates were needed the cone index should be used. However one problem with this conclusion as outlined by Ossola et al. (2015) is that even less literature have explored the variability of urban soil properties with empirical measurements. The study conducted by Ossola et al. (2015) explored soil permeability for three different habitat types, low-complexity parks which are minimally disturbed, high-complexity parks which were not actively managed and high-complexity remnants which represented the natural vegetation of the area. They concluded that soil permeability differed depending on the habitat type.

It is clear from the literature that simply assuming a flat percentage of runoff across catchments is wrong because different urban areas generate different values. Even within a singular catchment runoff values can vary. Alongside the fact that urban soil properties can differ from catchment to catchment and very depending on the soil type.

2.3.2 Hydrological urban routing methods

Similar to runoff generation, the process of routing in a catchment is influenced by urbanisation through the introduction of storm water drainage systems to transport runoff out of the built-up areas faster than natural processes. Similar to urban runoff generation, no one routing method has so far been proven to be the best. The two main methods of routing can be defined as a hydrological or hydraulic approach (Guo, 2006, p. 437). Hydrological routing uses simple methods such as continuity equations or storage relationships, to establish a relationship between inflow and outflow at specific points (Guo, 2006, p. 437). Hydrological routing methods include: linear/nonlinear Muskingum method (Linsley et al., 1949), kinematic wave routing (Lighthill and Whitham, 1955), and the Muskingum-Cunge (Cunge, 1969).

In contrast the hydraulic routing approach are based on the Saint-Venant equations, in which flow is determined as an inflow outflow response through numerical integration of flow equations across space and time (Mays and Tung, 2002, p. 411). This method is more data intensive than the hydrological routing, requiring data such as cross-sectional channel geometry, river bed and bank characteristics (Saint-Venant, 1871).

Price et al. (1980) applied a Muskingum-Cunge equation to urban routing by applying the equation as an alternative simpler method to the existing lumped pipe routing method in the Wallingford model (Kidd and Lowing, 1979). The Wallingford model package contains four various models, (i) a peak flow model, (ii) a peak flow model incorporating pipe-slope optimisation, (iii) a hydrograph model for design and simulation, (iv) a hydrograph simulation model including full solution for surcharged flow. The existing semi-distributed routing method was a below-ground hydraulic model incorporating both Muskingum-Cunge and solutions to simultaneous differential equations, to model the full pipe network of catchments. The simpler lumped methodology applied assumes the routing as a singular pipeline lumping a complex pipe system into a single "branch" splitting the pipe up along the way into n smaller pipe sections. As pointed out by Price et al. (1980) the simpler lumped method does not require access to detailed data, just measurements of pipe systems. Results show that the lumped method generated good results, but needed further research into generalisations.

The Muskingum-Cunge methodology is implemented into the CITY DRAIN model (Achleitner et al., 2007) which can model different parts of the urban drainage system, (catchment, sewer system, storage devises, receiving water etc). The Muskingum approach is used in the model for either a combined (lumped) or separate system (semi-distributed) in order to model the flow of water through the catchment to the outlet. Another model which uses Muskingum-Cunge is the Watershed Bounded Network Model WBNM (Boyd et al., 1996); a semi-distributed model which can be used for flood studies, runoff from impervious and pervious catchment surfaces and flood routing through storage reservoir. This model can use three different methods to model flood routing in urban catchments: (i) a nonlinear lag routing method (Askew, 1970), (ii) Muskingum routing, or (iii) a simple time delay method (which lags the flow at the outlet by an arbitrary value).

The study conducted by Basnayaka and Sarukkalige (2011) used two modelling approaches to simulate runoff from an urban catchment utilising sub-daily data. The catchment was split into sub-catchments and each assigned manhole covers, which were modelled through linking the stormwater pipe network, applying the Laurenson method for surface runoff routing. The Laurenson method splits catchments into sub-catchments then uses the Muskingum procedure to rout through each sub-catchment through a non-linear storage (Laurenson, 1964). The second method was a hydraulic approach, modelling surface runoff using the fully-distributed XPSWMM model. The XPSWMM uses a 2 dimensional approach which splits the catchment into a 6m x 6m grid and models flow between each grid. Results showed that both models were capable of representing the urban catchment, but it was noted that the poor topographical data reduced the accuracy of the hydraulic approach. The results also showed that once calibrated the hydrological model improved more than the hydraulic model. Rehman et al. (2003) explored the difference between two different models when utilising sub-daily data. The first model used RAFTS (Goyen et al., 1991) a semi-distributed model which splits the catchment into smaller sub-catchments in order to model and then combines each flow through a simple routing method. The second model SOBEK (Hydraulics, 2005) is a fully distributed grid-based 2-dimensional hydraulic model commonly used in floodplain modelling. The comparison was based on data from two catchments. The results showed whilst model performance was similar, if detailed catchment data was available for calibration, and then the SOBEK model was preferred. The conclusion by Rehman et al. (2003) was that even if detailed hydrological data was not available the SOBEK model was still preferred due to the various information within the catchment it could provide. Whilst this conclusion is true for exploring flow within a catchment Rehman et al. (2003) do not discus if no spatially varying data is available which model is preferred for lumped catchment outlet modelling. They noted that catchment response is very complex and difficult to model even with fully distributed models. It is argued in the literature that urban routing on a daily time step is potentially ineffective. For example Mitchell et al. (2001) concluded that excess rainfall will have left most catchments in a matter of hours. Hence flow routing in urban areas is unlikely to be effective on a daily time scale. Conversely, Figure 3 in Salvadore et al. (2015) showed the spatial and temporal scales of hydrological processes from a number of various studies, indicating that sewers and storm drainage lag-time are over a day long for catchments over 100 km^2.

Summary of Section

Similar to Section 2.3.1 a review of the literature show that there is no consensus on how best to route runoff through urban areas as all of the methods are constrained by data availability or require detailed urban pipeline characteristics/ measurements. It is generally acknowledged that with sufficiently detailed catchment data a distributed hydraulic model performs the best. However questions are raised concerning generalisation of the techniques as advanced data may not be readily available for every catchment. Hence simpler methods can be applied to catchments where only very little data is present. The next section will explore the issue of when complexity within models become redundant.

2.3.3 Complexity of models

An important issue with modelling, in general, is the complexity of models, such as how many model processes are needed before adding more become redundant. As outlined by Jacobson (2011) applying more detailed data to a model will lead to better performance, but there is a limitation to the level of detail of data that can be applied to every model, such that additional data will not improve model performance. For example, a complex model may require spatial data and fractional land-use data. If this data is not available, then model performance is expected to be lower and in some cases the model cannot be applied. However a simpler model cannot apply complex data and the detailed data would not improve model performance. This is a conclusion which is shared with Grayson et al. (2002) such that applying simple data to a complex model does not generate as good results than if more detailed data is applied. This conclusion is supported by Jakeman and Hornberger (1993) who found that the accuracy of model parameters can even be limited dependent on the amount and type of data present.

One methodology in order to reduce the impact of varying amounts of data for multiple catchments is by applying a multi-model structure. For example Butts et al. (2004) compared ten different model structures of the MIKE SHE and MIKE 11 models. Their results showed large variations in model performance and suggest that using a combination of model results rather than a single version of any one model would be preferable. The results also showed that increasing model complexity by adding additional processes did not improve model performance in some instances. They stated that future research is needed in evaluating the performance of model structure ensembles and advocate the use of a multiple model structure.

There is currently no definitive answer to the question of what constitutes optimal model complexity. One methodology is the multi model approach presented above. Another approach can be taken by considering a nested model structure. A nested model is a model which is a subset of another more complex model. For example, all of model A's parameters and processes are included into model B, but model B has more processes. In this case model A is a nested model. The advantage of this methodology is that if certain data is not available then the simpler nested model can be applied. Dominion (1999) states that for ecological system, simple models are a starting point and should be built upon into a more detailed model. These two idea (multi-model and nested model structure) will be combined for this thesis, creating a simple nested model structure, which can account for urban runoff and routing processes, but can be built upon. This creates research question 1 from Section 1.2

Research question 1: Can rainfall-runoff modelling in urban catchments be improved via implementation of a conceptual urban framework onto an already existing rural based model?

This research question will be explored further in Chapter 5: Impact of time-steps on rainfall runoff models performance in urban catchments.

Whilst the conclusion from Section 2.2.4 indicated that no particular model structure (lumped, semi-distributed and fully distributed) was the best, the studies presented compared models on predominately rural catchments. The analysis presented in Section 2.3.2 again indicated that no particular routing structure (lumped or semi-distributed) was preferred. Coupled with the analysis presented by Atkinson et al. (2003) who compared eight versions of a similar model structure when applied using both daily and sub-daily data, on a catchment in New Zealand. They concluded that the most advanced fully distributed version of the model performed the best. However, it was noted that some of the less detailed models did produce results of similar accuracy of the most detailed model. Conversely Dotto et al. (2011) compared two models: (i) a semi-distributed model (MUSIC) and (ii) a lumped model (KAREN), in order to explore parameter sensitivity. The study was conducted on five urban catchments in Melbourne, Australia with sub-daily data. With the conclusion that both models performed similarly and were able to reproduce the observed data. So Hypothesis 2 from Chapter 1.2 will explore these arguments further utilising the model created from research question 1.

Research question 2: What is the loss of information between lumped and spatially distributed rainfall-runoff models on a varying time scale for urban catchments.

This research question will be explored further in Chapter 6: A comparison between lumped and semi-distributed models utilising sub-daily data on the Gyeongancheon and Rodbourne catchments..

2.4 Part 4: Model calibration and Validation

This section will review current methods for calibration and validation of rainfall-runoff models. Model parameter calibration is defined as the process of adjusting a set of model parameters to ensure the best possible reproduction response of reality between observed data and model prediction within an acceptable range of specified accuracy Refsgaard and Henriksen (2004). Model parameter calibration is important, so that a model can appropriately represent the conditions of a catchment in order to generate accurate simulations. Model validation as defined by Sargent (2013), is "substantiation that a model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model."

There are two main approaches to parameter calibration for hydrological models. The first is conducting physical field experiments in order to estimate the parameters. The second approach is through use of an automatic global search algorithm to identify the set of parameters resulting in the closest fit between observed and simulated runoff. In order to apply such method a suitable calibration and validation period needs to be selected. Klemeš (1986) proposed four methods of model calibration and validation: (i) Split-sample test, (ii) Proxy-basin test, (iii) Differential split-sample test and (iv) Proxy-basin differential split-sample test.

The split-sample test is conducted by splitting the available data into two non-overlapping periods; data from the first period is then used for calibration, and data from the other period used for validation. This method is widely used by the hydrological community, e.g. Refsgaard (1997), Ewen (2011), and Donnelly-Makowecki and Moore (1999). However, all four methods have been scrutinised by a number of authors over the recent years. Mroczkowski et al. (1997) stated that simply applying the split sample-test should be the minimum requirement when testing model performance, and that the split-sample test is not an adequate test of comparing model structures. This study used a hypothesis style approach, with multi response data in the form of stream flow, stream chloride and groundwater levels data. The conclusion was that simply applying a split-sample test using stream flow data has limited explanatory power and that, whilst more advanced methods can reject models, the more advanced methods are needed to further model comparison tools. Andréassian et al. (2009) reviewed the four methods by Klemeš, presenting results from Le Moine (2008) which indicate that applying the proxy-basin, differential split-sample and proxy-basin differential methods can create a drop in model performance compared to the split-sample test. Hence, Andréassian et al. (2009) concluded that hydrological modellers prefer to just use the simpler split-sample method rather than the other three methods, arguing that due to this modellers may not fully consider what could be learnt from applying the other three calibration methods and just see the drop in model performance. This position was echoed by Seibert (2003) who concluded that models are typically not calibrated using Klemeš proxy-basin, differential split-sample, and proxy-basin differential methods to avoid having to deal with a decrease in model performance. Seibert (2003) argued the differential method presented by Klemeš (1986) such that models should be calibrated on time periods with dissimilar hydrological conditions, was a more powerful method but noted that only a few studies actually use it. Similarly Kirchner (2006) state that simply applying the split-sample test is inadequate because the calibration and validation time periods could have similar hydrological conditions, suggesting the split-sample tests conducted when hydrological conditions are different. Kirchner (2006) concluded that whilst better models need to be developed better analytical tools are also needed.

All of these studies agree that in addition to a simple split-sample test, method (ii)-(iv) of Klemeš (1986) list or more advanced method should be applied when assessing the performance of rainfall-runoff models. Whilst the split-sample method remains popular and generally accepted the literature suggests that more advanced methods can be employed to gain additional insight into model performance. As highlighted by a number of studies these methods are not being routinely used in the hydrological literature, hence simpler and more operational methods may need to be devised. But as stated in Refsgaard et al. (2005) developing new model calibration and validation tests has been neglected since Klemeš (1986). They argue that further development on testing schemes should be a major future challenge. Recently, advancements have been made in order to improve upon calibration and validation tests. For example, Ewen and O'Donnell (2012) explored the split-sample test further by splitting the period into three sections instead of two: a first calibration period, a second calibration period, and a validation period. The second calibration period is for calibrating the interval in which the model parameters perform the best. They acknowledge that while the method was successful in calibrating and validating a model on the 260 km^2 river Hodder catchment, further extensive testing is needed for a wide variety of catchments and storm responses in order for the method to be recommended for more general use. Gharari et al. (2013) also further explored model calibration by splitting the calibration period into multiple shorter calibration periods in order to obtain multiple parameter sets from which the single best set can be selected. Whilst this method is arguably a more robust calibration method due to more parameter sets being generated it does not attempt further model validation which is more important than simply generating multiple parameter sets. In summary the literature indicates that there are multiple methods to calibrate and validate models however these appear not to have found wide spread acceptance and use by the hydrological community. Hence, there is a need to develop new calibration and validation methodologies which can be easily applied to any model that requires parameter calibration. This concept of creating a new simple calibration and validation methodology forms part of research question 3 from Chapter 1.2.

Research question 3: Can simple statistically based techniques be used to improve validation of rainfall-runoff models. The idea of creating a new calibration and validation method will be explored further in Chapter 4: Operational model comparison techniques for rainfall-runoff models. This method is designed in order for the statistical based techniques from research question 3 can be applied.

2.5 Part 5: Model verification and selection techniques

Once suitable calibration and validation periods have been selected, model performance needs to be evaluated. Krause et al. (2005) outlined three key reasons as to why evaluating model performance is important: (i) To provide a quantitative estimate of the model's ability to reproduce historic and future watershed behaviour. (ii) To provide a means for evaluating improvements to the modelling approach through adjustment of model parameter values, model structural modifications, the inclusion of additional observational information, and representation of important spatial and temporal characteristics of the watershed. (iii) To compare current modelling efforts with previous study results. Current methods typically rely on comparing observed and simulated data via a performance criteria such as Nash-sutcliffe efficiency (NSE) shown in Equation 2.1:

$$\text{NSE} = 1 - \frac{\sum_{t=1}^{n}(q_{obs}(t) - q_{sim}(t))^2}{\sum_{t=1}^{n}(q_{obs}(t) - \bar{q}_{obs}(t))^2}. \tag{2.1}$$

Root Mean Square Error (RMSE) shown in Equation 2.2;

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n}(q_{obs}(t) - q_{sim}(t))^2}{n}}, \tag{2.2}$$

or the coefficient of determination (R^2):

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(q_{obs}(t) - \hat{q_{sim}}(t))^2}{\sum_{t=1}^{n}(q_{obs}(t) - \bar{q}_{obs}(t))^2}. \tag{2.3}$$

For each of the three equations presented and the rest to be presented within Section 2.5, $q_{obs}(t)$ is observed flow, $q_{sim}(t)$ is model simulated flow, $\hat{q_{sim}}(t)$ is predicted values from a statistical model and $\bar{q}_{obs}(t)$ is the mean of the observed flow. A performance criteria is defined by Beven (2011) as a mathematical measures of how well a model simulation fits available observations. An alternative name to performance criteria used within the literature is efficiency criteria. These two terms can be used interchangeably but, for the purposes of this thesis the term performance criteria will be used.

However there are a number of issues associated with performance criteria that need to be addressed. The first problem is which performance criteria to choose Krause et al. (2005) compared nine different performance criteria, concluding that no single performance criteria can be considered the best. Each criteria has advantages and disadvantages associated with them, highlighting that different criteria are sensitive to different parts of river flow, and that selection of criteria should be based upon the intended use of the model. Similarly Weglarczyk (1998) concluded that using multiple performance criteria can be misleading due to the inter-dependence between them, warning that if multiple criteria are to be used then care needs to be taken not to draw contradictory conclusions.

The second problem associated with performance criteria is that performance criteria can be inaccurate. Willmott and Matsuura (2005) highlight problems with RMSE as it is a function of three steps: (i) the total square error, (ii) division by n number of data points, and (iii) square root. This has the implication that large errors introduced in the first step will only become increasingly large through the other steps, resulting in a larger negative impact. Whilst an absolute value method such as the mean absolute error (MAE), shown in Equation 2.4 does not have a tendency to inflate large errors and as such was preferred by Willmott and Matsuura (2005):

$$\text{MAE} = \frac{\sum_{t=1}^{n}|q_{obs} - q_{sim}|}{n}. \tag{2.4}$$

Gupta et al. (2009) argued that the Nash-Sutcliffe efficiency (NSE) is sensitive to periods of larger observed runoff values. This is because evaluation of NSE involves squaring the differences, and thus larger runoff values create larger differences if the model under or over predicts observed runoff. This conclusion is also supported by Legates and McCabe (1999) and Krause et al. (2005). McCuen et al. (2006) also evaluated the NSE concluding the criteria is sensitive to a number of factors, such as sample size and outliers due to the NSE being a singular value which can be inaccurate over longer periods of time.

In order to overcome some of these problems alternative performance criteria have been developed such as the Kling-Gupta efficiency (KGE), which is an equal weighting of three components: (i)correlation, (ii) bias, (iii) and variability measures (Gupta et al., 2009) as shown in Equation 2.5 when σ denotes the respected standard deviation:

$$\mathrm{KGE} = 1 - \sqrt{(CC-1)^2 + \left(\frac{\sigma_{q_{sim}}}{\sigma_{q_{obs}}} - 1\right)^2 + \left(\frac{q_{\bar{sim}}}{q_{\bar{obs}}} - 1\right)^2} \tag{2.5}$$

$$\mathrm{CC} = \frac{\frac{1}{n}\sum_{t=1}^{n}\left(q_{sim} x q_{obs}\right) - q_{\bar{obs}} . q_{\bar{sim}}}{\sigma_{q_{obs}} . \sigma_{q_{sim}}}. \tag{2.6}$$

Whilst results indicate this is a better measure than the NSE, the main conclusion by Gupta et al. (2009) is that no matter which performance criteria is used one of the most important parts of modelling remains model calibration.

Whilst new performance criteria can be designed to overcome the inaccuracies of performance criteria, the issue of how to interpret performance criteria still remains. This issue was raised by Criss and Winston (2008) who highlighted that the problem with NSE is that the lower bound is negative infinity but a negate value of the NSE does not necessarily indicate poor model performance. As the NSE is based on the mean of the observed flow, so if the flow is steady then the denominator of the NSE is small, resulting in a larger negative value. In order to solve the issue of large negative values Criss and Winston (2008) propose a new efficiency criteria called the volumetric efficiency (VE), which similar to the KGE, does not involve squaring of differences between observed and simulated runoff. But unlike the KGE, the VE has a lower bound of zero.

McCuen et al. (2006) argue that a problem with performance criteria is that it is just a singular value and therefore cannot adequately capture variation in simulations. Jain and Sudheer (2008) argue that one of the most important problems with analysing performance criteria is that 5% over estimation of observed flow will likely have different implications from a 5% under estimation, which could indicate zero flow. Thus, simply reporting a performance criteria value does not indicate over or under estimation as it is a singular lumped value. Similarly Legates and McCabe (1999) compare a number of performance criteria, concluding that performance criteria can be misleading due to its sensitivity to extreme values, and thus simply applying and stating performance criteria is too simple and as a minimum the reasoning behind the choice of criteria needs to be made clear. Kirchner et al. (1996) state that, as a minimum, model performance evaluation should have three key elements: (i) a performance criterion, (ii) a benchmark model, in which the tested model is being tested against, and (iii) an outcome, detailed by how much better or worse the tested model does against the benchmark model. Similarly Jain and Sudheer (2008) discusses the disadvantages of using the NSE and show that a poor model was able to obtain a high NSE value, so the use of other tools when examining model performance is needed. This result is shared by a number of studies such as Schaefli and Gupta (2007) who concluded that simply relying on the Nash-Sutcliffe efficiency alone is not sufficient to validate a model. They suggest when applying performance criteria to determine model performance a simple benchmark model should also be applied such that this benchmark model fits the basic requirement for the case study in order to provide evidence that the model being tested is sufficiently good.

Whilst performance criteria are the more commonly used method of model evaluation, alternative methods are applied which do not take performance criteria into account. For example, Hooper (2001) argued that applying performance criteria is not sufficient to validate a model, and instead a hypothesis style approach should be adopted, in which the model is tested multiple times using different criteria to obtain its limitations. However, the main problem with this approach is that whilst it does provide a method to critique a model, catchments are unique and the conclusions drawn from one catchment will not necessarily apply to other catchments. Krause et al. (2005) criticise performance criteria arguing that it is too simple to analyse models and in their argument state that using observed and simulated stream flow hydro graphs in order to strengthen the comparison argument.

A method proposed by Vogel and Sankarasubramanian (2003) using the idea that model performance should be performed prior to and independently of parameter calibration, therefore removing calibration error. The methodology presented is a generalized sensitivity analysis which is employed to analyse the performance of models on certain key hydrological characteristics of observed and simulated data. The authors also state that when applying performance techniques they should reflect on the goals of the modelling excise and purpose of the hydrological model.

Studies have incorporated the aforementioned methods into their analysis, such as Bouffard (2014) who compared two models and used both performance criteria and uncertainty analysis. Anh et al. (2010) compared three models using three different performance criteria and graphical analysis on a single catchment, with conclusions that each of the models performance was good and that the choice of model would be dependent on the purpose of the study. Similarly Refsgaard and Knudsen (1996) used a combination of graphical and performance criteria to compare three models of varying complexity on three different catchments. Boyle et al. (2001) compared two models, one lumped and one semi-distributed on a Blue river watershed located in Oklahoma, in which hydrographs of observed and model simulations are used to analyse specific river flow events. Whilst all of these studies have applied various techniques in order to strengthen results, they all still apply Performance criteria. Thus, methods need to be developed in order to better incorporate Performance criteria into analysis. One such method is presented in Ewen (2011) who present a dynamic programming algorithm in order to create visual performance criteria which measure differences in amplitude and timing errors between simulated and observed hydrographs. This method is described by the authors as an extension to performance criteria.

In summary, it is clear from the literature that performance criteria are widely used in operational and procedural hydrology but used unconditionally can be a misleading tool. There are already more advanced methodologies used in order to compare hydrological models but these methods are rarely used. As such simple methods need to be developed in order to incorporate these techniques into hydrological modelling. Which results in hypothesis 3 from Chapter 1.2:

Hypothesis 3: Can simple statistical based techniques be used to improve validation of rainfall-runoff models.

This hypothesis will be explored further in Chapter 4: Operational model comparison techniques for rainfall-runoff models.

Chapter 3

Model development

This chapter details the development process of the URMOD model, starting with a description and derivations of the underlying model (DAYMOD) representing the non-urban (or rural) catchment hydrology. Next, details of the new urban infiltration models (Section 3.2.1) and routing extensions (Section 3.2.5) are presented followed by details of model calibration and validation (Section 3.2.6). These sections are followed by a comparison between the three proposed urban infiltration models (Section 3.3), with a final best framework presented.

3.1 Rural Model processes- DAYMOD

As outlined in Chapter 1, research question 1 was to investigate if an urban framework can be implemented onto an existing model which does not account for urban processes, thus creating a nested model structure. The rural model selected as the base for the nested urban model structure is the rural model DAYMOD, which was first developed for the Antecedent Rainfall Project with the aim to improve the Flood Estimation Handbook (FEH) rainfall-runoff models. The original description of DAYMOD in Packman (2004) featured a comparison between eight lumped soil moisture accounting models, coupled with a singular baseflow and surface flow routing method. The DAYMOD model featured in this thesis is the best performing soil moisture model from the report alongside the routing methodology presented in the report.

DAYMOD is a lumped conceptual parameter-parsimonious model with seven calibrated parameters representing two main processes: (i) infiltration and runoff generation and (ii) baseflow and channel routing which is represented in Figure 3-1. Infiltration and runoff generation occurs when precipitation that falls onto the ground either infiltrates into the soil (left-hand-side of Figure 3-1) or becomes direct runoff (right-hand-side of Figure 3-1). Direct runoff (or excess rainfall) is the rainfall that does not infiltrate into the ground and thus is transferred into runoff. The precipitation that infiltrates into the soil is handled by a conceptual soil column based approach, with the contribution lost to evaporation or drainage out of the column conditional on soil moisture levels. The soil moisture is calculated as a result of infiltration, evaporation and drainage. Whilst the remaining infiltrated precipitation is retained in the soil column as soil moisture.

The precipitation that does not infiltrate into the soil becomes runoff, which is then split again with a portion contributing to the baseflow whilst the remaining runoff contributes to the surface flow. Baseflow is the section of flow that sustains the river between rainfall events, the baseflow is led to the river via delayed pathways. Whilst in real world processes baseflow has a contribution from infiltrated precipitation, DAYMOD does not take this into account and baseflow is only made up of non infiltrated precipitation. The drainage as shown on Figure 3-1 drains out of the soil column and the precipitation is lost. The surface flow is defined as the contribution of the river from overland flow. The routing model in DAYMOD, routs the baseflow through two linear reservoirs, a baseflow reservoir with a time delay constant (B_L) and then the channel reservoir with a shorter delay (S_L). Whilst the surface flow is only routed through the channel reservoir.

DAYMOD has a number of parameters which need to be to be calibrated, requiring the following observed data: (i) observed rainfall, (ii) potential evaporation and (iii) observed river flow. Originally DAYMOD was calibrated using a simplex method for function minimisation as outlined in Nelder and Mead (1965). However the shuffled evolution complex algorithm of Duan et al. (1993) was adapted in this thesis. Further details on how model calibration is performed is outlined in Section 3.2.6. Once DAYMOD has been calibrated and a set of parameters is obtained, then a time series of rainfall (q_{obs}) and potential evaporation (E_p) data can be used as input to obtain DAYMODs outputs which are catchment average soil moisture, baseflow (b_t), surface flow (s_t) and actual evaporation (E_a). Figure 3-1 is a visual representation of the hydrological water cycle that DAYMOD represents.

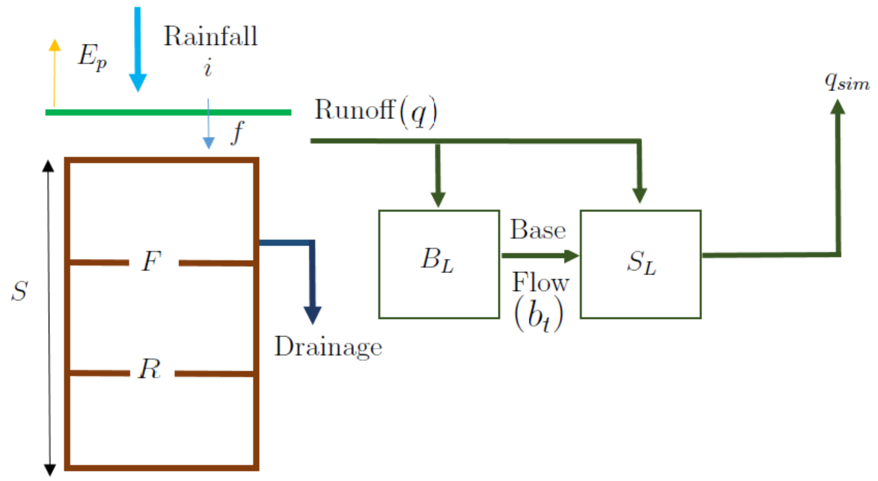


Figure 3-1: Full model diagram of DAYMOD, with conceptual soil column and zones on the left hand side. Routing system with time delay boxes on the right.

For reference, Table 3.1 and 3.2 has a list of all input, output and parameter notation for DAYMOD, and the urban model URMOD. The following subsections will describe how infiltration and runoff is handled for the original DAYMOD, with the final section detailing the routing method used.

Parameter	Description	Unit
S	Average Soil moisture capacity	mm
F	Field capacity	mm
R	Rooting Depth	mm
k	Drainage coefficient	-
Φ	Proportion of runoff split	-
B_L	Baseflow lag	day
S_L	Channel lag	day
U_L	Urban lag	day
γ	Scaling term	-

Table 3.1: URMOD model parameters, description and units

Input	abbreviation	Unit
Observed rainfall	i	mm
Observed riverflow	q_{obs}	m^3/s
Potential evaporation	E_p	mm
Ouput	abbreviation	Unit
Simulated riverflow	q_{sim}	m^3/s
Simulated baseflow	b_t	m^3/s
Average soil moisture	-	mm
Actual evaporation	E_a	mm

Table 3.2: URMOD inputs and outputs

3.1.1 Infiltration and runoff generation

Infiltration and runoff generation in DAYMOD is based on a conceptual soil column approach such that soil moisture level (m) of the column for a time step (t) is made up of three contributions: (i) infiltration, (ii) drainage and (iii) evaporation. The model is based on the premise that precipitation that does not infiltrate into the ground becomes runoff. The percentage of infiltration is handled by a infiltration factor denoted (f) which can be written as a contribution of the runoff (κ). Similarly the runoff factor denoted (q) can be written as a contribution of the infiltration (η), due to a simple mass-balance rain (i) = infiltration (η) + runoff (κ) as shown in Equation 3.1:

$$\kappa = i(1-f) \quad \text{or} \quad \eta = i(1-q). \tag{3.1}$$

To determine the infiltration (η) it is first necessary to determine the runoff factor (q). The runoff model used in DAYMOD is based on the uniform Probability Distributed Model, (PDM), Moore (1985). The PDM assumes that the catchment is split into a number of individual storages. The soil moisture capacity of these storages is (C) and the size varies uniformly over the entire catchment between a value of zero and C_{max}. The distribution of C values are random over the catchment, but is assumed to be statistically uniform, hence each capacity occurs with equal frequency. This approach is commonly used within hydrology (Moore and Bell (2002) and Moore (1985)) and within other models such as the

Xinanjiang model (Ren-Jun, 1992) and the Arno model (Todini, 1996). The assumption is shown in (Sivapalan and Woods, 1995) who within the paper explore the cumulative frequency of soil depths. If all the soil capacities of the catchment are ordered in size a geometrical representation would form a triangle as shown in Figure 3-2. The probability of a storage being smaller than or equal to C_{max} is 1, as shown in Equation 3.4

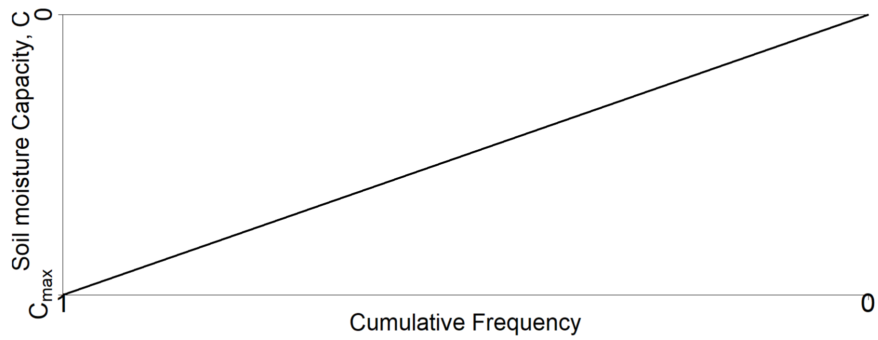


Figure 3-2: Geometric representation of cumulative catchment soil column capacity

From Figure 3-2 the probability density function $f(C)$ is given by;

$$f(C) = \frac{1}{C_{max}}, \qquad 0 \leq C \leq C_{max} \quad 0 \text{ elsewhere,} \tag{3.2}$$

distribution function $F(c)$ can then be generated

$$F(C) = \int_0^C \frac{1}{C_{max}} dt = \frac{C}{C_{max}}. \tag{3.3}$$

At C_{max} the cumulative distribution is:

$$F(C_{max}) = \frac{C_{max}}{C_{max}} = 1. \tag{3.4}$$

If an initial level of soil moisture is assumed (C_0), then any areas with a capacity less than C_0 are saturated while areas with soil moisture capacity excess of C_0 are unsaturated with a deficit ($C - C_0$). Next, consider a precipitation event (i). This added water will further reduce the deficit in the unsaturated areas to ($C - C_0 - i$), while 100% runoff is assumed from the saturated areas. This is depicted in Figure 3-3.

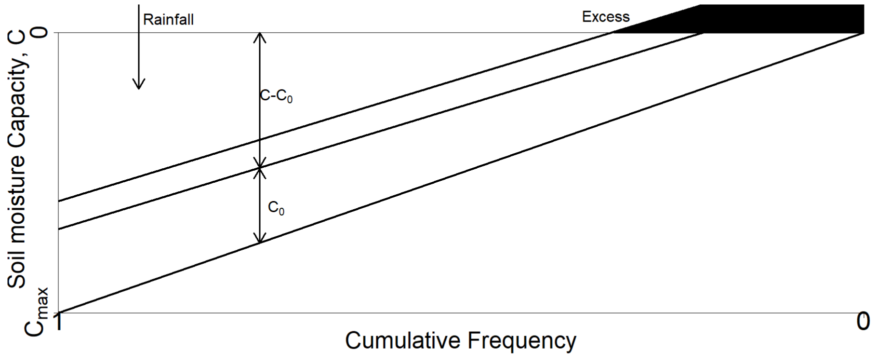


Figure 3-3: Representation of catchment runoff generation and soil saturation. C_0 denotes the amount of soil moisture that is saturated, C is the soil moisture capacity and excess is the rainfall that did not infiltrate into the soil.

Since DAYMOD is a continuous model after ($t = 0$) the soil moisture content (m) will vary through out time, dependent on rainfall. Due to DAYMOD being a lumped representation of a catchment in order to model the change in soil moisture a average soil moisture capacity across the entire catchment (S) needs to be defined. From Figure 3-3 the average soil moisture capacity (S) in a catchment can be determined as:

$$S = \frac{1}{2}C_{max}. \tag{3.5}$$

From geometrical consideration of Figure 3-3, initially at time $t = 0$, the proportion of the catchment that is unsaturated is $\frac{(C_{max}-C_0)}{C_{max}}$ and the unsaturated ranges uniformly from zero to $C_{max} - C_0$. The unsaturated volume of the catchment is defined as:

$$S_{un} = \frac{1}{2}(C_{max} - C_0)\frac{(C_{max} - C_0)}{C_{max}}. \tag{3.6}$$

The mean actual moisture content m_0 at $t = 0$ can be defined as the mean soil moisture capacity minus the mean unsaturated volume. Hence the mean actual moisture content can be defined as:

$$m_0 = \frac{1}{2}C_{max} - \frac{\frac{1}{2}(C_{max} - C_0)^2}{C_{max}}. \tag{3.7}$$

So the average catchment moisture content can be defined as $\frac{m_0}{S}$, which is a normalised mean between [0,1]:

$$\frac{m_0}{S} = \frac{0.5C_{max} - \frac{0.5(C_{max}-C_0)^2}{C_{max}}}{0.5C_{max}}. \tag{3.8}$$

Which can be simplified:

$$\frac{m_0}{S} = 1 - \left(\frac{0.5\left(C_{max} - C_0\right)^2}{0.5(C_{max})^2}\right) \tag{3.9}$$

$$\frac{m_0}{S} = 1 - \left(\frac{(C_{max})^2 - 2C_0C_{max} + (C_0)^2}{(C_{max})^2}\right) \tag{3.10}$$

$$\frac{m_0}{S} = 1 - \left(1 - \frac{C_0}{C_{max}}\right)^2. \tag{3.11}$$

From Figure 3-3 the proportion of precipitation that is turned into runoff is $\frac{C_0}{C_{max}}$. As outlined in Equation 3.1 rainfall $i = \eta + \kappa$, precipitation that has not infiltrated is turned into runoff, hence $\left(1 - \frac{C_0}{C_{max}}\right)$ is the corresponding catchment average percentage infiltration. So rearranging Equation 3.11 to isolate $\left(1 - \frac{C_0}{C_max}\right)$:

$$\frac{m_0}{S} = 1 - \left(1 - \frac{C_0}{C_{max}}\right)^2 \tag{3.12}$$

$$\left(1 - \frac{m_0}{S}\right) = \left(1 - \frac{C_0}{C_{max}}\right)^2 \tag{3.13}$$

$$\left(1 - \frac{m_0}{S}\right)^{\frac{1}{2}} = \left(1 - \frac{C_0}{C_{max}}\right). \tag{3.14}$$

Thus $\left(1 - \frac{m_0}{S}\right)^{\frac{1}{2}}$ is the catchment average percentage infiltration. Dropping the suffix gives the general soil moisture content such that the infiltration factor f can be defined as:

$$f = \left(1 - \frac{m}{S}\right)^{\frac{1}{2}}. \tag{3.15}$$

Figure 3-4 shows varying values of f (black line) as a percentage as a function of relative soil moisture m.

As can be seen from Figure 3-4 the ratio $\frac{m}{S}$ is defined between zero and

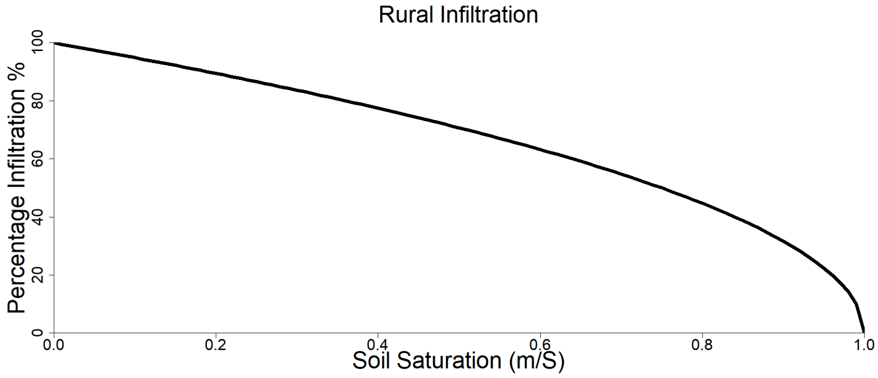


Figure 3-4: Percentage infiltration ($f \times 100$) versus soil saturation $\frac{m}{S}$ for rural runoff.

one, zero indicating that the soil column is completely dry where as a value of one indicates that the soil column is completely saturated. As outlined in the introduction of this chapter, alongside the infiltration factor, drainage and evaporation govern the amount of moisture in the soil column.

In order to determine the drainage and evaporation terms, the structure of the soil column is analysed. The soil column for DAYMOD takes two soil restrictions into account: (i) field capacity and (ii) rooting depth. Field capacity (F) is defined as the amount of soil moisture retained in the soil column after drainage. When soil moisture is above field capacity, drainage and evaporation remove water from the soil column. Below field capacity drainage ceases, but evaporation still occurs. Whereas in some models field capacity is a physically measured parameter, DAYMOD is a conceptual model so a direct physical interpretation is not possible, hence F is a calibrated parameter. Rooting depth (R) is defined as the soil moisture level at which the roots of vegetation reach. When soil moisture is above rooting depth but below field capacity there is evaporation, but below rooting depth evaporation is reduced linearly with depth. Whilst rooting depths can vary depending on the vegetation (i.e large trees and small plants have very different length roots), only a single calibrated general parameter is used in the model. Figure 3-5 shows the soil column from Figure 3-1. As can be seen from Figure 3-5 there are three zones split by the two half lines representing rooting depth and field capacity.

In order to determine how the soil moisture is affected in each zone, a soil moisture equation needs to be constructed. Since the soil moisture behaves dif-

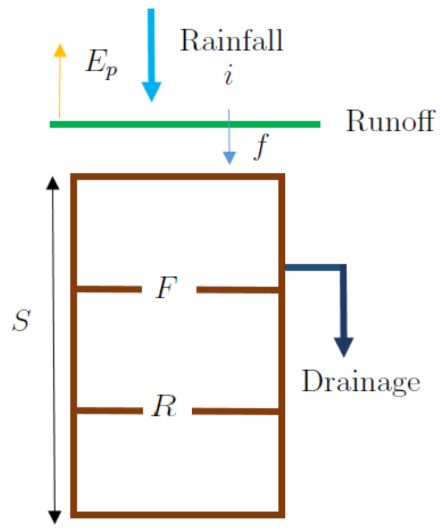


Figure 3-5: Conceptual soil column from DAYMOD Figure 3-1, with rainfall, evaporation, and soil column processes.

ferently in each of the three zones, three different soil moisture equations are needed. Zone one is when moisture content is above field capacity ($m > F$). As such drainage and evaporation happens, with evaporation loss assumed to happen at the potential rate (E_p), whilst drainage depends upon moisture content (m) and a calibrated drainage coefficient (k) so drainage out of the soil column takes place at a rate of $k(m - F)$. Combining the infiltration factor, drainage and evaporation gives the change in soil moisture equation at a given time t as shown in Equation 3.16:

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \underbrace{i\left(1 - \frac{m}{S}\right)^{\frac{1}{2}}}_{\text{Infiltration}} - \underbrace{k(m - F)}_{\text{Drainage}} - \underbrace{E_p}_{\text{Evaporation}} . \tag{3.16}$$

Zone 2 is when the soil moisture is between the field capacity and the rooting depth ($R < m < F$). As stated previously drainage ceases, whereas evaporation loss happens at the potential rate (E_p), so the soil moisture equation at a given time t can be presented as:

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \underbrace{i\left(1 - \frac{m}{S}\right)^{\frac{1}{2}}}_{\text{Infiltration}} - \underbrace{E_p}_{\text{Evaporation}} . \tag{3.17}$$

Zone 3 is when soil moisture is below the rooting depth ($m < R$). In zone 3 there is again no drainage but evaporation reduces linearly with depth $E_p \frac{m}{R}$. Equation 3.18 is the change in soil moisture equation for this zone:

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \underbrace{i\left(1 - \frac{m}{S}\right)^{\frac{1}{2}}}_{\text{Infiltration}} - \underbrace{E_p \frac{m}{R}}_{\text{Evaporation}} . \tag{3.18}$$

Equation 3.16, 3.17 and 3.18 are solved so the direct runoff value can be determined. Simple analytical solutions for Equations 3.16, 3.17 and 3.18 do not exist hence a finite difference method has to be applied. The finite difference method is a numerical method used to solve differential equations through approximation of differentials over a time step Δt. Whilst alternative numerical methods can be applied to solve the equations, the finite difference method will be used as it is the same method used originally by Packman (2004). For the full derivations see Packman (2004). Equation 3.16 solution is;

$$M_t = G - \left(\frac{i_*}{k_*}\right)^2 + \left(\frac{i_*}{k_*}\right)\sqrt{\left(\frac{i_*}{k_*}\right)^2 - 2G + 4\left(1 - \frac{M_0}{2}\right)}, \tag{3.19}$$

with the following substitutions:

$$M_t = \frac{m_t}{S}, \quad i_* = \frac{i\Delta t}{2S}, \quad k_* = 1 + \frac{k\Delta t}{2}, \quad E_* = \Delta t \frac{E_p - kF}{S} \tag{3.20}$$

$$G = M_0 \frac{2 - k_*}{k_*} - \frac{E_*}{k_*}. \tag{3.21}$$

Equation 3.17 solution is;

$$M_t = G - (i_*)^2 + (i_*)\sqrt{(i_*)^2 - 2G + 4\left(1 - \frac{M_0}{2}\right)}, \tag{3.22}$$

with the following substitutions:

$$M_t = \frac{m_t}{S}, \quad i_* = \frac{i\Delta t}{2S}, \quad E_* = \Delta t \frac{E_p}{S} \tag{3.23}$$

$$G = M_0 - E_*. \tag{3.24}$$

Equation 3.18 solution is;

$$M_t = G - \left(\frac{i_*}{k_*}\right)^2 + \left(\frac{i_*}{k_*}\right)\sqrt{\left(\frac{i_*}{k_*}\right)^2 - 2G + 4\left(1 - \frac{M_0}{2}\right)}, \tag{3.25}$$

with the following substitutions:

$$M_t = \frac{m_t}{S}, \quad i_* = \frac{i\Delta t}{2S}, \quad k_* = 1 + \frac{\Delta t E_p}{2R} \tag{3.26}$$

$$G = M_0 \frac{2 - k_*}{k_*}. \tag{3.27}$$

In order to determine the soil moisture at a certain time step the mean over the timestep t is taken:

$$\overline{m} = \frac{m_0 + m_t}{2}. \tag{3.28}$$

But when the moisture m crosses a zone mid timestep, the timestep is divided and the remaining period uses the new zone. The next section describes how runoff is treated.

3.1.2 Routing, baseflow and surface flow generation

As described in Equation 3.1 precipitation that does not infiltrate into the soil becomes runoff. So runoff κ can be defined as one minus the infiltration factor as presented in Equation 3.29:

$$\kappa = i(1 - f) = i\left(1 - \left(1 - \frac{m}{S}\right)^{\frac{1}{2}}\right). \tag{3.29}$$

As can be seen from Equation 3.29 runoff generation is not assumed to be a linear process and depends upon soil saturation (m/s). If the soil is completely saturated $(m/s = 1)$ runoff generation is equal to the amount of rainfall i. In contrast, if the soil is completely dry $(m/s = 0)$ then the runoff generation equals zero. Figure 3-6 below shows the routing section from Figure 3-1.

The precipitation designated as runoff is split into slow responding baseflow and fast responding surface flow. These two distinct flow components are then combined at the outlet to form total runoff. The proportion split between the

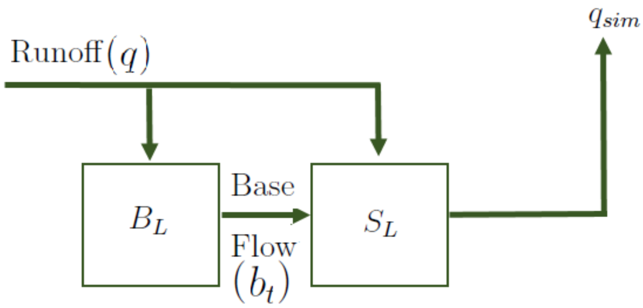


Figure 3-6: Representation of routing in DAYMOD, with 2 linear reservoirs and combined flow outlet.

two types of flow is controlled by a calibrated parameter denoted ϕ, such that ϕ is the proportion for baseflow, whereas $(1 - \phi)$ is the proportion of the surface flow contribution. As shown in Figure 3-6 the runoff that is designated as baseflow is first routed through a local linear reservoir (local baseflow reservoir) with a time constant of B_L before it emerges into the river channel where it is subsequently routed to the catchment outlet via a linear reservoir (local surface flow reservoir) with time constant S_L. The surface flow is only routed through the channel reservoir i.e, bypassing the local linear reservoir. The routing method for DAYMOD is based on the linear reservoir concept, with a characteristic recession defined as an exponential decay. The baseflow recharge (r) that feeds the local baseflow reservoir, with storage σ_B is given as B_L times the outflow b. As shown in Equation 3.30

$$\sigma_B = B_L b. \tag{3.30}$$

The change in storage for a timestep t can be determined as:

$$\frac{\mathrm{d}\sigma_B}{\mathrm{d}t} = r - b. \tag{3.31}$$

The continuity equation Equation 3.31 can be combined with Equation 3.30 to give a simple first order linear differential equation:

$$B_L \frac{\mathrm{d}b}{\mathrm{d}t} + b = r. \tag{3.32}$$

The linear reservoir equation can then be multiplied by $e^{\frac{t}{B_L}}/B_L$ and rearranged

$$\frac{\mathrm{d}\left(be^{\frac{t}{B_L}}\right)}{\mathrm{d}t} = \left(\frac{r}{B_L}\right)e^{\frac{t}{B_L}}. \tag{3.33}$$

Equation 3.33 can be integrated over a time step $\Delta t = t - t_0$ between 0 and t which gives the local baseflow rate Equation 3.34

$$b_t = b_0 e^{\frac{-t}{B_L}} + r\left(1 - e^{\frac{-t}{B_L}}\right). \tag{3.34}$$

The mean baseflow for a time step t can then be obtained by averaging flow over the time step $(b_0 + b_t)/2$, this is then routed though the channel routing model. The average of b_t is taken if the observed data is averaged, where as if instantaneous data is obtained then averaging is not needed. The derivation for the channel routing model uses the same methodology as for the baseflow model. The surface flow recharge is denoted z, the storage σ_S is given as S_L times the outflow s, as shown in Equation 3.35

$$\sigma_S = S_L s. \tag{3.35}$$

Following a similar method method for Equation 3.34 the channel routing model is obtained in Equation 3.36

$$s_t = s_0 e^{\frac{-t}{S_L}} + z\left(1 - e^{\frac{-t}{S_L}}\right). \tag{3.36}$$

Once the mean baseflow is routed through the channel routing model (Equation 3.36), the mean is then obtained again which results in the baseflow at the catchment outlet for a time step t. Unlike the baseflow, the direct surface flow is the runoff designated to be surface flow is only routed through Eq 3.36, such that $(s_0 + s_t)/2$ is the average surface flow at the catchment outlet. Combining the baseflow and surface flow gives the total flow at the catchment outlet (q_{sim}). In summary, this section has presented the two main processes of the rural model DAYMOD. The next section will detail the urban frameworks which were implemented onto the rural sections of the model, Section 3.2.1 will detail the urban runoff frameworks, whereas Section 3.2.5 will detail the urban routing framework.

3.2 Urban Model processes- URMOD

Five urban frameworks are presented in this section, three urban runoff generation frameworks are presented in Section 3.2.1 and two urban routing frameworks are presented in Section 3.2.5.

3.2.1 Urban infiltration generation

This section will detail the new urban infiltration and runoff generation framework developed in this thesis. In the default rural model (DAYMOD) the infiltration factor (f) is described by a single lumped term Equation 3.15. For the urban extension the infiltration has been split into two distinct sources: one representing infiltration in urban areas, whilst the other represents infiltration in the rural areas of the catchment. Note that urban areas are a combination of impervious (e.g. streets) and pervious areas (e.g. parks). Due to this infiltration occurs within the pervious areas of urban areas. The resulting model is shown in Equation 3.37;

$$f = i(1-u)f_{rur} + iuf_{urb}, \tag{3.37}$$

where f_{rur} represents infiltration in the rural areas, f_{urb} denotes infiltration in urban areas, and u represents the fraction of the catchment covered by urban surfaces. So if $u = 0$ then the infiltration factor would default back to Equation 3.15 and describing an entirely rural catchment. The infiltration factor for rural areas (f_{rur}) is the same as in DAYMOD hence:

$$f_{rur} = \left(1 - \frac{m}{S}\right)^{\frac{1}{2}}. \tag{3.38}$$

Whereas if $u = 1$ this would imply that the area was completely urbanised and so only the urban infiltration factor would be used. Next, the infiltration and runoff generation across the urban areas will be considered. Three urban extension methods are proposed in order to account for impervious surfaces in catchments. The first method assumes a fixed percentage of rainfall infiltrates and is turned into runoff for urban areas decoupled from soil moisture. Review of the literature in Section 2.3.1 indicated that the value changes depending on the type of impervious surface. The value chosen for this thesis is 70% as suggested in Kjeldsen (2009). The second method attempts to account for the infiltration of the pervious sections of the urban areas, rather than assuming a fixed percentage runoff as presented in the first method. In contrast to the fixed percentage in this method the infiltration depends on soil moisture, and at a certain threshold of soil moisture, 100% of rainfall is translated into runoff from impervious areas. This method attempts to incorporate conceptual hydrological processes into the runoff generation, such that urban areas do not generate a fixed amount of runoff but do generate more runoff than rural areas. The third method assumes that runoff and infiltration generation in the urban areas are dependent on a calibrated scaling term denoted γ, such that runoff generation from the urban and rural areas are similar but γ denotes the percentage reduction in infiltration across the urban areas. This method attempts to account for the variability in infiltration in the urban areas. If $\gamma = 0$ then infiltration for the urban area is the same as the rural area and whilst $\gamma = 1$ assumes that the urban area is completely sealed as such no infiltration would be present.

3.2.2 Urban infiltration framework 1: Fixed percentage runoff

Urban framework 1 assumes a fixed percentage runoff from the urban area of a catchment, which is independent of the soil moisture and rainfall until a certain m/S value. This fixed percentage runoff will be denoted ω_{imp}. Comparing ω_{imp} to the runoff generated for the rural parts, it is clear that if ω_{imp} is less than 100 % there will be a soil moisture threshold when the percentage runoff generated from the rural (pervious) areas is greater than that of the impervious areas, i.e :

$$1 - \left(1 - \frac{m_0}{S}\right)^{\frac{1}{2}} > \omega_{urb} \Rightarrow \frac{m_0}{S} > 1 - (1 - \omega_{imp})^2. \tag{3.39}$$

For example, if $\omega_{imp} = 70\%$ then the threshold value of soil moisture is $1-(1-0.70)^2) = 0.91$. Therefore infiltration from the urban areas has to be considered for soil moisture levels both above and below this threshold level. If the soil moisture level exceeds this threshold then the urban areas will revert to a function like the rural areas, i.e:

$$f_{urb} = \begin{cases} 1-\omega_{imp} & : \frac{m}{S} \leq 1-(1-\omega_{imp})^2 \\ (1-\frac{m}{S})^{\frac{1}{2}} & : \frac{m}{S} > 1-(1-\omega_{imp})^2. \end{cases} \tag{3.40}$$

Figure 3-7 shows the urban infiltration f_{urb}, alongside the rural infiltration, for the case when $\omega_{imp} = 0.7$

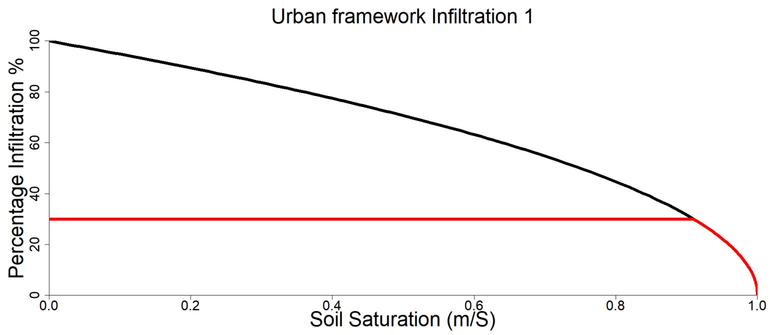


Figure 3-7: Percentage infiltration versus soil saturation. Urban framework 1: Urban infiltration (Red), Rural infiltration (Black)

By substituting Equation 3.40 into Equation 3.37 the total infiltration factors accounting for both the pervious and impervious areas can be defined as

$$f = \begin{cases} (1-u)(1-\frac{m}{S})^{\frac{1}{2}} + iu(1-\omega_{imp}) & : \frac{m}{S} \leq 1-(1-\omega_{imp})^2 \\ (1-\frac{m}{S})^{\frac{1}{2}} & : \frac{m}{S} > 1-(1-\omega_{imp})^2. \end{cases} \tag{3.41}$$

The consequence of introducing the infiltration model in Equation 3.41 is that when soil moisture is high ($m/S > 1-(1-\omega_{imp})^2$) the soil moisture equation defaults to the rural model DAYMOD. This is due to the rural areas soil being saturated, so the runoff generated is more than that of the urban areas. Which results in the effect of urbanisation begin reduced. The soil moisture accounting equations and derivations for high soil moisture are the same as DAYMOD (Equation 3.16, 3.17 and 3.18) and will not be derived again. However the soil moisture levels below ($m/S \leq 1-(1-\omega_{imp})^2$) a set of three new soil moisture equations need to be constructed to account for the change in infiltration across urban areas; one for each of three zones of the soil moisture column described in Section 3.1.1. The consequence of changing the δt would be rather than a daily time step it would be sub-daily. This will mean the soil moisture will change zones less often due to the number of time steps increasing. First, Equation 3.42 is for zone 1 when moisture is above field capacity, ($m > F$):

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \underbrace{i\left((1-u)\left(1-\frac{m}{S}\right)^{\frac{1}{2}} + u(1-\omega_{imp})\right)}_{\text{infiltration}} - \underbrace{k(m-F)}_{\text{Drainage}} - \underbrace{E_p}_{\text{Evaporation}} . \quad (3.42)$$

Equation 3.43 is for zone 2 when moisture is between field capacity and the rooting depth ($R < m < F$):

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \underbrace{i\left((1-u)\left(1-\frac{m}{S}\right)^{\frac{1}{2}} + u(1-\omega_{imp})\right)}_{\text{infiltration}} - \underbrace{E_p}_{\text{Evaporation}} . \quad (3.43)$$

Zone 3, when soil moisture is below the rooting depth ($m < R$), Equation 3.44 is the change in soil moisture equation for this zone:

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \underbrace{i\left((1-u)\left(1-\frac{m}{S}\right)^{\frac{1}{2}} + u(1-\omega_{imp})\right)}_{\text{infiltration}} - \underbrace{E_p\frac{m}{R}}_{\text{Evaporation}} . \quad (3.44)$$

Similar to DAYMOD a finite difference method is needed in order to solve Equation 3.42, 3.43 and 3.44. Unlike the rural derivations, the full derivation for Equation 3.42 will be presented whereas condensed derivations will be presented for Equation 3.43 and 3.44. Taking a finite difference for Equation 3.42 gives:

$$\frac{m_t - m_0}{t - t_0} = i\left((1-u)\left(1-\frac{m_t+m_0}{2S}\right)^{\frac{1}{2}} + u(1-\omega_{imp})\right) - k\left(\frac{m_t+m_0}{2} - F\right) - E_p. \quad (3.45)$$

Rearranging for the infiltration term gives

$$i\Delta t\left(1-\frac{m_t+m_0}{2S}\right)^{\frac{1}{2}}(1-u) = (m_t-m_0) - i\Delta t u(1-\omega_{imp}) + k\Delta t\left(\frac{m_t+m_0}{2} - F\right) + \Delta t E_p. \quad (3.46)$$

Squaring and dividing by S^2 gives

$$\left(\frac{i\Delta t}{S}\right)^2\left(1-\frac{m_t+m_0}{2}\right)(1-u)^2 =$$
$$\left(\frac{m_t}{S}\left(1+\Delta t\frac{k}{2}\right)-\frac{m_0}{S}\left(1-\Delta t\frac{k}{2}\right)+\Delta t\frac{E_p-kF-iu(1-\omega_{imp})}{S}\right)^2. \quad (3.47)$$

Making the following substitutions, and dividing by k_*

$$M_t=\frac{m_t}{S},\quad i_*=\frac{i\Delta t}{2S},\quad k_*=1+\frac{k\Delta t}{2},\quad E_*=\Delta t\frac{E_p-kF-iu(1-\omega_{imp})}{S}. \quad (3.48)$$

Noting that 2-k_* =1-$\frac{k\Delta t}{2}$, gives

$$\left(\frac{2i_*}{k_*}\right)^2\left(1-\frac{M_t+M_0}{2}\right)(1-u)^2=\left(M_t-M_0\frac{2-k_*}{k_*}+\frac{E_*}{k_*}\right)^2. \quad (3.49)$$

A new constant G_u can be defined as

$$G_u=M_0\frac{2-k_*}{k_*}+\frac{E_*}{k_*}. \quad (3.50)$$

Resulting in

$$\left(\frac{2i_*}{k_*}\right)^2\left(1-\frac{M_t+M_0}{2}\right)(1-u)^2=\left(M_t-G_u\right)^2. \quad (3.51)$$

The second-order polynomial in M_t is given below

$$\underbrace{M_t^2}_{a}+\underbrace{\left(\left(\frac{2i_*}{k_*}\right)^2\frac{(1-u)^2}{2}-2G_u\right)M_t}_{b}+\underbrace{\left(G^2-(1-u)^2\left(\frac{2i_*}{k_*}\right)^2\left(1-\frac{M_0}{2}\right)\right)}_{c}=0. \quad (3.52)$$

Which has solution

$$M_t=G_u-(1-u)^2\left(\frac{i_*}{k_*}\right)^2+(1-u)\left(\frac{i_*}{k_*}\right)\sqrt{(1-u)^2\left(\frac{i_*}{k_*}\right)^2-2G_u+4\left(1-\frac{M_0}{2}\right)}. \quad (3.53)$$

The same method is applied to Equations 3.43 and 3.44 in order to determine there respective solutions. The full derivations are not presented only the solutions.

Equation 3.43 has solution:

$$M_t = G_u - (1-u)^2(i_*)^2 + (1-u)(i_*)\sqrt{(1-u)^2(i_*)^2 - 2G_u + 4\left(1 - \frac{M_0}{2}\right)}, \quad (3.54)$$

with the following substitutions

$$M_t = \frac{m_t}{S}, \quad i_* = \frac{i\Delta t}{2S}, \quad E_* = \Delta t \frac{E_p - iu(1-\omega_{imp})}{S}, \quad G_u = M_0 + E_*. \quad (3.55)$$

Whilst Equation 3.44 has solution:

$$M_t = G_u - (1-u)^2\left(\frac{i_*}{k_*}\right)^2 + (1-u)\left(\frac{i_*}{k_*}\right)\sqrt{(1-u)^2\left(\frac{i_*}{k_*}\right)^2 - 2G_u + 4\left(1 - \frac{M_0}{2}\right)}, \quad (3.56)$$

with the following substitutions

$$M_t = \frac{m_t}{S}, \quad i_* = \frac{i\Delta t}{2S}, \quad k_* = 1 + \frac{\Delta t E_p}{R}, \quad G_u = M_0\frac{2-k_*}{k_*} - \frac{i\Delta t u(1-\omega_{imp})}{k_*}. \quad (3.57)$$

Whilst this is the same solution as Equation 3.53, the G_u and E_p terms are different. Such that there is no evaporation in Equation 3.57. As stated in the introduction of Section 3.2.1, a default value of 70% is chosen for ω_{imp}. Alterations can be made such that ω_{imp} could be a calibrated term, whilst this would make the urban model more flexible and in-line with current literature as outlined in Chapter 2.3.1 such that runoff from impervious surfaces is not a fixed value and changes depending on the urban surfaces. However the purpose of Urban framework 1 was to test a fixed value and a calibrated term for urban infiltration is presented in the third framework.

3.2.3 Urban infiltration framework 2: Combined increasing and fixed urban runoff

Unlike urban framework 1, runoff generation in urban framework 2 depends on soil moisture, such that due to urban areas having less pervious surfaces than rural areas, the soil column will fill up faster and reach saturation quicker than that of rural areas. Upon reaching saturation runoff from urban areas will be 100%, hence 0% infiltration. When the soil column is completely dry, the runoff generation from the urban areas will still be larger than that of the rural areas. This value will be denoted ω_{imp} and similar to urban framework 1 will be chosen as 70%. The amount of runoff generated will increase until 100% runoff is reached and will continue until soil moisture drops below the threshold value. This process is shown in Figure 3-8.

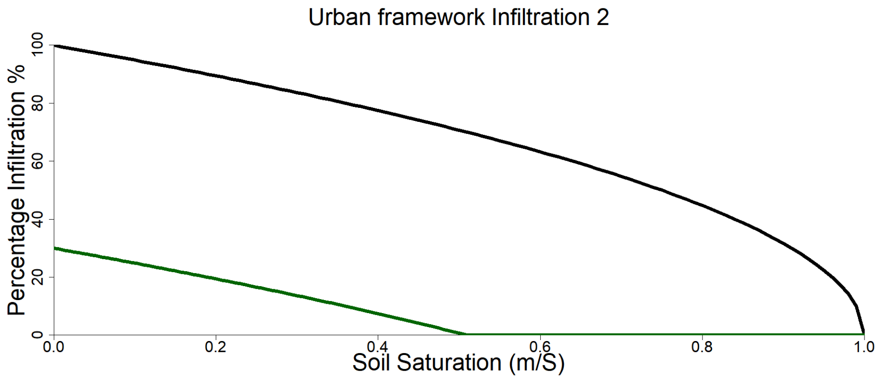


Figure 3-8: Percentage infiltration versus soil saturation. Urban framework 2: Urban infiltration (Dark green), Rural infiltration (Black)

The two urban infiltration terms can be written as:

$$f_{urb} = \begin{cases} i\left(1 - \frac{m}{S}\right)^{\frac{1}{2}} - i\omega_{imp} & : \frac{m}{S} \leq 1 - \omega_{imp}^2 \\ 0 & : \frac{m}{S} > 1 - \omega_{imp}^2. \end{cases} \tag{3.58}$$

For the low soil moisture (m/S$\in [0, 1 - \omega_{imp}^2]$) the infiltration equation is:

$$f = i(1-u)f_{rur} + iuf_{urb} \tag{3.59}$$

$$f = i(1-u)\left(1-\frac{m}{S}\right)^{\frac{1}{2}} + iu\left(1-\frac{m}{S}\right)^{\frac{1}{2}} - u\omega_{imp} \tag{3.60}$$

$$f = i\left(1-\frac{m}{S}\right)^{\frac{1}{2}} - iu\left(1-\frac{m}{S}\right)^{\frac{1}{2}} + iu\left(1-\frac{m}{S}\right)^{\frac{1}{2}} - iu\omega_{imp} \tag{3.61}$$

$$f = i\left(1-\frac{m}{S}\right)^{\frac{1}{2}} - iu\omega_{imp}. \tag{3.62}$$

Whereas for the high soil moisture ($m/S \in [1-\omega_{imp}^2, 1)$) the infiltration consists solely of infiltration from the rural area, i.e the infiltration equation becomes:

$$f = i(1-u)f_{rur} + iuf_{urb} \tag{3.63}$$

$$f = i(1-u)\left(1-\frac{m}{S}\right)^{\frac{1}{2}} + 0. \tag{3.64}$$

So the total infiltration for both cases is given as:

$$f = \begin{cases} i(1-\frac{m}{S})^{\frac{1}{2}} - iu\omega_{imp} & : \frac{m}{S} \leq 1-\omega_{imp}^2 \\ i(1-\frac{m}{S})^{\frac{1}{2}}(1-u) & : \frac{m}{S} > 1-\omega_{imp}^2. \end{cases} \tag{3.65}$$

There are three soil moisture equations depending on the zone location of the moisture in the soil column, however for urban framework 2, there are two cases for each zone, when soil moisture is high or low. For low soil moisture ($m/S \leq 1-\omega_{imp}^2$) the moisture equation for moisture above field capacity is expressed as the equation below:

$$\frac{\mathrm{d}m}{\mathrm{d}t} = i\left(\left(1-\frac{m}{S}\right)^{\frac{1}{2}} - u\omega_{imp}\right) - k(m-F) - E_p. \tag{3.66}$$

The second case is high soil moisture when ($m/S > 1-\omega_{imp}^2$). The equation for when high soil moisture and when moisture is above field capacity:

$$\frac{\mathrm{d}m}{\mathrm{d}t} = i(1-u)\left(1-\frac{m}{S}\right)^{\frac{1}{2}} - k(m-F) - E_p. \tag{3.67}$$

Similar to the previous derivation only the final solutions and substitutions will be presented. Hence the solution to Equation 3.66 is:

$$M_t = G_u - \left(\frac{i_*}{k_*}\right)^2 + \left(\frac{i_*}{k_*}\right)\sqrt{\left(\frac{i_*}{k_*}\right)^2 - 2G_u + 4\left(1 - \frac{M_0}{2}\right)}. \tag{3.68}$$

With the following substitutions

$\mathrm{M}_t = \frac{m_t}{S}, \quad i_* = \frac{i\Delta t}{2S}, \quad k_* = 1 + \frac{k\Delta t}{2}, \quad E_* = \Delta t\frac{E_p - kF + iu\omega_{imp}}{S}$
$G_u = M_0\frac{2-k_*}{k_*} + \frac{E_*}{k_*}$

The case when soil moisture is high $\left(\frac{m}{S} > 1 - \omega_{imp}^2\right)$ is presented:

$$M_t = G_u - (1-u)^2\left(\frac{i_*}{k_*}\right)^2 + (1-u)\left(\frac{i_*}{k_*}\right)\sqrt{(1-u)^2\left(\frac{i_*}{k_*}\right)^2 - 2G_u + 4\left(1 - \frac{M_0}{2}\right)}. \tag{3.69}$$

With the following substitutions

$$M_t = \frac{m_t}{S}, \quad i_* = \frac{i\Delta t}{2S}, \quad k_* = 1 + \frac{k\Delta t}{2}, \quad E_* = \Delta t\frac{E_p - kF}{S}, \quad G_u = M_0\frac{2-k_*}{k_*} + \frac{E_*}{k_*} \tag{3.70}$$

The soil moisture equation's and solutions for when the soil moisture is between rooting depth and field capacity ($R < m < F$), are now presented for both low and high soil moisture. The low soil moisture,

$$\frac{\mathrm{d}m}{\mathrm{d}t} = i\left(\left(1 - \frac{m}{S}\right)^{\frac{1}{2}} - u\omega_{imp}\right) - E_p \tag{3.71}$$

Whereas the high soil moisture equation is

$$\frac{\mathrm{d}m}{\mathrm{d}t} = i(1-u)\left(1 - \frac{m}{S}\right)^{\frac{1}{2}} - E_p \tag{3.72}$$

The solution to the low soil moisture equation (Equation 3.72) will now be presented:

$$M_t = G_u - (i_*)^2 + (i_*)\sqrt{(i_*)^2 - 2G_u + 4\left(1 - \frac{M_0}{2}\right)} \tag{3.73}$$

Making the following substitutions

$$M_t = \frac{m_t}{S}, \quad i_* = \frac{i\Delta t}{2S}, \quad E_* = \Delta t \frac{E_p + iu\omega_{imp}}{S}, \quad G_u = M_0 - E_p \tag{3.74}$$

The final solution to the high soil moisture equation (Equation 3.73) is

$$M_t = G_u - (1-u)^2(i_*)^2 + (1-u)(i_*)\sqrt{(1-u)^2(i_*)^2 - 2G_u + 4\left(1 - \frac{M_0}{2}\right)} \tag{3.75}$$

With the following substitutions

$$M_t = \frac{m_t}{S}, \quad i_* = \frac{i\Delta t}{2S}, \quad E_* = \Delta t \frac{E_p}{S}, \quad G_u = M_0 - E_* \tag{3.76}$$

The final two equations to be solved are the low and high soil moisture for when soil moisture is below the rooting depth ($m < R$). Equation 3.78 is the case for the low soil moisture

$$\frac{\mathrm{d}m}{\mathrm{d}t} = i\left(\left(1 - \frac{m}{S}\right)^{\frac{1}{2}} - u\omega_{imp}\right) - E_p\frac{m}{R} \tag{3.77}$$

And for the high soil moisture:

$$\frac{\mathrm{d}m}{\mathrm{d}t} = i(1-u)\left(1 - \frac{m}{S}\right)^{\frac{1}{2}} - E_p\frac{m}{R}. \tag{3.78}$$

The solution to 3.79 is:

$$M_t = G_u - \left(\frac{i_*}{k_*}\right)^2 + \left(\frac{i_*}{k_*}\sqrt{\left(\frac{i_*}{k_*}\right)^2 - 2G_u + 4\left(1 - \frac{M_0}{2}\right)}\right) \tag{3.79}$$

With the following substitutions

$$M_t = \frac{m_t}{S}, \quad i_* = \frac{i\Delta t}{2S}, \quad k_* = 1 + \frac{\Delta t E_p}{2R}, \quad G_u = M_0\frac{2 - k_*}{k_*} - \frac{i\Delta t u\omega_{imp}}{Sk_*} \tag{3.80}$$

The solution to 3.79 is presented:

$$M_t = G_u - (1-u)^2 \left(\frac{i_*}{k_*}\right)^2 + (1-u)\left(\frac{i_*}{k_*}\right)\sqrt{(1-u)^2\left(\frac{i_*}{k_*}\right)^2 - 2G_u + 4\left(1 - \frac{M_0}{2}\right)} \tag{3.81}$$

With the following substitutions

$$M_t = \frac{m_t}{S}, \quad i_* = \frac{i\Delta t}{2S}, \quad k_* = 1 + \frac{\Delta t E_p}{2R}, \quad G_u = M_0 \frac{2 - k_*}{k_*} \tag{3.82}$$

This methodology is more flexible than urban framework 1 as runoff generation is not fixed and is dependent on the soil moisture from the soil column. Alongside the fact, urban framework 2 attempts to incorporate real world ideas into the model, such that increasing runoff during wetter periods due to urban surfaces sealing from precipitation.

3.2.4 Urban infiltration framework 3- Soil moisture dependent increasing urban runoff

Urban framework 3 assumes infiltration across the urban areas is dependent on soil moisture, but is reduced by a factor γ.

$$f_{urb} = i(1-\gamma)\left(1 - \frac{m}{S}\right)^{\frac{1}{2}}. \tag{3.83}$$

The γ term attempts to account for the variability in infiltration in different urban areas as outlined in Section 2.3.1, such that a large γ indicates that the urban area is mostly impervious and more runoff is generated, whilst a smaller value indicates that the urban area has more pervious surfaces and more runoff is generated. If γ is zero then infiltration for the impervious area is the same as the infiltration for the rural area, whereas if γ is one then the impervious area would be sealed and there would be no infiltration. Figure 3-9 shows the infiltration for the urban framework 3 f_{urb}, this is the case when $\gamma = 0.7$

The total infiltration factor is expressed below,

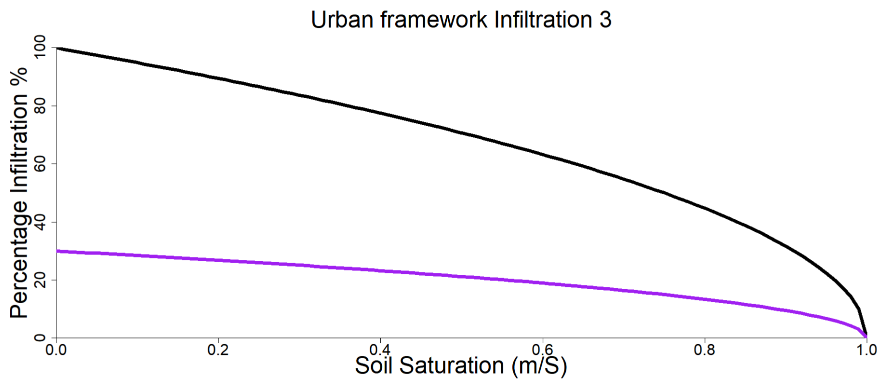


Figure 3-9: Percentage infiltration versus soil saturation. Urban framework 3: Urban infiltration (purple), Rural infiltration (Black)

$$f = i(1-u)\left(1-\frac{m}{S}\right)^{\frac{1}{2}} + iu(1-\gamma)\left(1-\frac{m}{S}\right)^{\frac{1}{2}} \tag{3.84}$$

$$f = i(1-u\gamma)\left(1-\frac{m}{S}\right)^{\frac{1}{2}} \tag{3.85}$$

hence the soil moisture equation for moisture above field capacity ($m > F$) is defined as

$$\frac{\mathrm{d}m}{\mathrm{d}t} = i(1-u\gamma)\left(1-\frac{m}{S}\right)^{\frac{1}{2}} - k(m-F) - E_p. \tag{3.86}$$

The soil moisture equation for moisture between field capacity and rooting depth ($R < m < F$)

$$\frac{\mathrm{d}m}{\mathrm{d}t} = i(1-u\gamma)\left(1-\frac{m}{S}\right)^{\frac{1}{2}} - E_p. \tag{3.87}$$

The final zone, when soil moisture is below the rooting depth ($m < R$)

$$\frac{\mathrm{d}m}{\mathrm{d}t} = i(1-u\gamma)\left(1-\frac{m}{S}\right)^{\frac{1}{2}} - E_p\frac{m}{R}. \tag{3.88}$$

Similar to the two previous sections the full derivation for the soil moisture equation will be omitted and only the solutions and substitutions are shown. The solution to Equation 3.87 is presented:

$M_t =$

$$G_u - (1-u\gamma)^2 \left(\frac{i_*}{k_*}\right)^2 + (1-u\gamma) \left(\frac{i_*}{k_*}\right) \sqrt{(1-u\gamma)^2 \left(\frac{i_*}{k_*}\right)^2 - 2G_u + 4\left(1 - \frac{M_0}{2}\right)} \tag{3.89}$$

Making the following substitutions

$$M_t = \frac{m_t}{S}, \quad i_* = \frac{i\Delta t}{2S}, \quad k_* = 1 + \frac{k\Delta t}{2}, \quad E_* = \Delta t \frac{E_p - kF}{S}, \quad G_u = M_0 \frac{2-k_*}{k_*} - \frac{E_p}{k_*} \tag{3.90}$$

The solution to Equation 3.88 is presented:

$$M_t = G_u - (1-u\gamma)^2 (i_*)^2 + (1-u\gamma)(i_*) \sqrt{(1-u\gamma)^2 (i_*)^2 - 2G_u + 4\left(1 - \frac{M_0}{2}\right)}. \tag{3.91}$$

Making the following substitutions

$$M_t = \frac{m_t}{S}, \quad i_* = \frac{i\Delta t}{2S}, \quad E_* = \frac{\Delta t E_p}{S}, \quad G_u = M_0 + E_* \tag{3.92}$$

The solution to Equation 3.89 is

$$M_t = G_u - (1-u\gamma) \left(\frac{i_*}{k_*}\right)^2 + (1-u\gamma) \left(\frac{i_*}{k_*}\right) \sqrt{(1-u\gamma)^2 \left(\frac{i_*}{k_*}\right)^2 - 2G_u + 4\left(1 - \frac{M_0}{2}\right)} \tag{3.93}$$

Making the following substitutions

$$M_t = \frac{m_t}{S}, \quad i_* = \frac{i\Delta t}{2S}, \quad k_* = 1 + \frac{\Delta t E_p}{2R}, \quad G_u = M_0 \frac{2-k_*}{k_*} \tag{3.94}$$

This method is the most flexible due to the scaling term γ being calibrated.

3.2.5 Urban routing methods

Once the direct runoff from the urban areas is calculated, it needs to be routed in order to determine the flow at the catchment outlet. Two routing methods have been developed for the urban models. The first is an extension of the routing method presented in Section 3.1.2 via a single linear reservoir, in which more surface flow is generated but the flow from the urban and rural areas are routed through the main channel with the same lag-time. The second method is a storm-drain pipe system approach, with an overflow like approach, such that if the storm-drain pipe system reaches full capacity, then excess direct runoff spill onto the rural part of the catchment and thus gets routed through the rural section. Similarly to the first method, more surface flow is routed but the urban runoff is routed faster than runoff from the rural areas. Unlike the routing model used in the rural model section which had only one contribution of direct runoff q, the urban model runoff is made up of two contributions the rural runoff q_{rur} and the urban runoff q_{urb}, such that $q = q_{rur} + q_{urb}$.

Urban routing method 1 - Single linear reservoir

The first urban routing method is an extension of the method presented in Section 3.1.2. Similar to the rural routing, the direct runoff from the rural model is split, with a contribution going to the baseflow, and the remaining contributing to the surface flow. The split in rural runoff is done via the ϕ parameter which was introduced in Section 3.1.2. The local baseflow equation is the same as in Section 3.1.2 as displayed in Equation 3.96:

$$b_t = b_0 e^{\frac{-t}{B_L}} + r(1 - e^{\frac{-t}{B_L}}), \tag{3.95}$$

where r is the rural runoff designated to be baseflow. Whereas in the rural channel routing model (Equation 3.36) there is one contribution (z) for surface runoff, for the urban extension there is two, the rural runoff contribution (z_r) and the urban runoff contribution(z_u). The urban runoff (z_u) does not contribute to the baseflow and is therefore considered to contribute directly to the surface flow. Due to two runoff contributions a new linear reservoir equation is needed for the surface flow, with storage denoted σ_3 such that the lag S_L times the total outflow u:

$$\sigma_u = S_L s. \tag{3.96}$$

The change in storage can therefore be defined as

$$\frac{\mathrm{d}\sigma_u}{dt} = z_r + z_u - s. \tag{3.97}$$

Substituting Equation 3.97 into Equation 3.98 gives

$$S_L \frac{\mathrm{d}v}{\mathrm{d}t} + s = z_r + z_u. \tag{3.98}$$

Multiplying through by $e^{\frac{t}{S_L}}/S_L$ gives

$$\frac{\mathrm{d}\left(s e^{\frac{t}{S_L}}\right)}{\mathrm{d}t} = \left(\frac{z_r}{S_L}\right) e^{\frac{t}{S_L}} + \left(\frac{z_u}{S_L}\right) e^{\frac{t}{S_L}}. \tag{3.99}$$

Using the same integration method presented in Section 3.1.2, over the time step 0 to t gives the downstream outflow rate

$$s_t = s_0 e^{\frac{-t}{S_L}} + z_r \left(1 - e^{\frac{-t}{S_L}}\right) + z_u \left(1 - e^{\frac{-t}{S_L}}\right). \tag{3.100}$$

The mean surface flow can be obtained by averaging over the time step $((s_0 + s_t)/2)$. As shown in Equation 3.101 both z_r and z_u get routed at the same speed, so the only difference from the routing method presented in Section 3.1.2 is a reduction in baseflow. Similar to Section 3.1.2 the baseflow and surface flow is combined to determine the outflow at the catchment outlet.

Urban routing method 2 - parallel linear reservoir

The second urban routing method is a bounded pipe-based approach, which has a overflow system such that if the pipe system reaches capacity then the excess flow contributes to the rural surface routing. The surface routing for the runoff from the rural areas is the same method as in the rural routing, as shown in Equation 3.102

$$s_t = s_0 e^{\frac{-t}{S_L}} + z_r \left(1 - e^{\frac{-t}{S_L}}\right). \tag{3.101}$$

In contrast to method 1, here the urban runoff z_u is routed via a parallel linear reservoir in the form of a bounded pipe-flow. The method is taken from (Motiee et al., 1997) which uses a hydraulic method to simulate a drainage network. A separate linear storage σ_p similar to Section 3.2.5 can be defined as Equation 3.103 noting that U_L is a calibrated urban lag-term and v is the outflow

$$.\sigma_p = U_L v \tag{3.102}$$

The change in storage σ_p can be defined as the difference between z_u and v

$$\frac{\mathrm{d}\sigma_p}{\mathrm{d}t} = z_u - v. \tag{3.103}$$

Combining equations 3.103 and 3.104 results in a first-order differential equation

$$U_L \frac{\mathrm{d}v}{\mathrm{d}t} + v = z_u. \tag{3.104}$$

Equation 3.105 can be solved using a finite difference method similar to the methodology used to solve the soil moisture equations in Section 3.2.4. For the derivation, taking a finite difference for Equation 3.105,

$$U_L \frac{v_t - v_0}{t - t_0} = \frac{z_t + z_0}{2} - \frac{v_t + v_0}{2}. \tag{3.105}$$

Rearranging to isolate the outflow term v_t

$$U_L(v_t - v_0) = \frac{\Delta t(z_t + z_0)}{2} - \frac{\Delta t(v_t + v_0)}{2}, \tag{3.106}$$

$$\frac{\Delta t(z_t + z_0)}{2} = U_L(v_t - v_0) + \frac{\Delta t(v_t + v_0)}{2}, \tag{3.107}$$

$$\frac{\Delta t(z_t + z_0)}{2} = v_t \left(U_L + \frac{\Delta t}{2}\right) - v_0 \left(U_L - \frac{\Delta t}{2}\right), \tag{3.108}$$

$$v_t \left(U_L + \frac{\Delta t}{2}\right) = \frac{\Delta t(z_t + z_0)}{2} + v_0 \left(U_L - \frac{\Delta t}{2}\right). \tag{3.109}$$

This can be simplified to

$$v_t = \frac{2U_L v_0 + \Delta t(z_t - v_0 + z_0)}{2(U_L + \Delta t)}. \tag{3.110}$$

As stated previously the pipe system is bounded such that if the capacity is reached the excess runoff contributes towards the rural model. Through testing which will not be presented within the thesis, it was shown that the upper bound and overflow did not have any significant impact on model performance due to the small amount of rainfall passing through the model system for catchments with limited rain. However when tested on a catchment with more rain (See Section 6) the overflow did have a slight impact on performance. As such the upper bound will be left in the model and set at 1 m^3/s, so further testing can be applied.

3.2.6 Model calibration

In order to determine the optimal model parameter set θ, a number of model parameters need to be calibrated. For the rural model, DAYMOD, seven parameters are in need of calibration: $(S, F, R, k, \phi, B_L, S_L) = \theta_r$. For the urban model frameworks two extra parameters need calibration: U_L, for the third urban model the soil-moisture scaling term γ.

Whilst Jakeman and Hornberger (1993) argue that a rainfall-runoff record can only detect 4-5 parameters, this model will keep all 7 rural and 2 urban parameters. This is because no changes to the initial model structure (DAYMOD) will be undertaken as this is a existing model and so changing the structure could hinder model performance. The purpose of this thesis is too develop a urban extension for existing rural models and not change them. Whilst the structure of the model will not be changed constraints are applied to the parameters in order to not obtain unreasonable values.

The set of URMOD parameters is denoted θ_u. Calibration of the model parameters is achieved via optimisation by finding the optimal set of model parameters θ minimising the value of an objective function using the shuffled complex evolution algorithm (SCE) Duan et al. (1993). The shuffled complex evolution algorithm samples from across the multidimensional space in order to determine the minimum of the objective function. However the SCE method does not converge initially, but rather samples from across the entire space in order to determine the parameters. The maximum number of iterations for the SCE method is chosen as 10000. This is because upon testing the model with less and more iterations it was found that no improvement in performance was obtained by increasing the number of iterations past 10000, whilst not as much time (less than 5 minutes for a 5 year calibration) was saved by decreasing the number of iterations. The model is run with an initial set of parameter values and observed calibration input data (rainfall, potential evaporation) to simulate the runoff at the catchment outlet. An initial set of parameters was chosen, the rural parameters were chosen from advice of the previous model user. For the U_L parameter, this value was chosen as half the initial value of the S_L, due to conceptually U_L should be smaller than S_L. The γ parameter was selected as 0.7, which would represent 70% similar to framework 1 and 2. The initial soil moisture is a fixed value dependent on the S parameter, whilst a calibrated initial soil moisture value was tested it was determined not to be implemented into the final model structure due to over parametrisation. Initial flow values are taken from observed data:

$$q_{sim} = M(\hat{\theta}|I). \tag{3.111}$$

The simulated runoff q_{sim} is then compared to the observed runoff for the calibration period q_{obs} in order to obtain the residual sum of squares:

$$error = \left| \sum_{i=1}^{n} (q_{obs} - q_{sim})^2 \right|. \tag{3.112}$$

Whilst for this section the residual sum of squares was chosen as the measure of performance, alternative measures can be chosen such as MAE or NSE. However, choosing a measure that has a squared value was preferred as outlined in Chapter 2 squared measures prioritise peak flows. The model then seeks to reduce the error (Equation 3.113) by optimising the parameters. This process continues for a maximum of 10000 iterations or Equation 3.113 reaches 0 i.e simulation perfectly matches the observation. Due to M being an approximate representation of the true catchment, Equation 3.113= 0 will never happen. Whilst the calibration process is running a set of constraints are enforced on the algorithm, this is so unrealistic parameters are not obtained such as a negative soil column size or a negative lag. The constraints applied to URMOD and DAYMOD are (i) No negative parameter values, i.e $\theta > 0$, (ii) the field capacity ratio must be greater than the routing depth ratio, but must remain less than 1. i.e $R \leq F \leq 1$, (iii) the drainage coefficient k must be greater than 0.1 but less than 1. $0.1 < k < 1$, (iv) the baseflow lag is larger than the channel lag and the channel lag is greater than zero. $B_L > S_L$, (v) the urban lag is smaller than that of the surface lag. $S_L > U_L$ and (vi) the urban scaling term γ must be greater than zero and less than one. $0 \leq \gamma \leq 1$. Once the parameter set obtained can then be used with rainfall and evaporation data to create simulated river runoff values and soil moisture for other periods of time.

The software all of the models was developed and coded in was R, R is a freely available open source program environment. The original DAYMOD code was checked to make sure no bugs were present by recreating older simulations of the model. Similarly when the urban models were designed and coded, at every step care was taken to ensure no bugs or coding errors were present. Currently the model takes approximately 15-20 minutes to perform a full calibration (5 years worth of data and 10000 iterations). This can change depending on the length of calibration period or if the number of iterations are increased or decreased.

3.3 Case Study- Comparison of urban runoff frameworks

All three urban infiltration model structures presented in Section 3.2, were applied to five catchments in order to compare their respective performance. The three urban infiltration models were then combined with the single linear reservoir routing method presented in Section 3.2.5. Five catchment were selected for initial model testing, these are summarised in Table 3.3, and based on the percentage of urbanisation catchment descriptor URBEXT_{2000} Bayliss et al. (2006). The National River Flow Archive (NRFA) classifies each of their gauging stations with a value. These values will be used as the name of the respective catchments.

NRFA	Name of catchment	Area (KM^2)	$URBEXT_{2000}$(%)
39033	Bagnor	49.2	0.12
39052	Binfield	50.2	23.71
39094	Marsh farm	96.67	49.17
39096	Wembley	21.8	50.8
39131	Costons Lane Greenford	146.2	52.82

Table 3.3: Table of catchments, NRFA gauging station values and name with size and percentage of urbanisation.

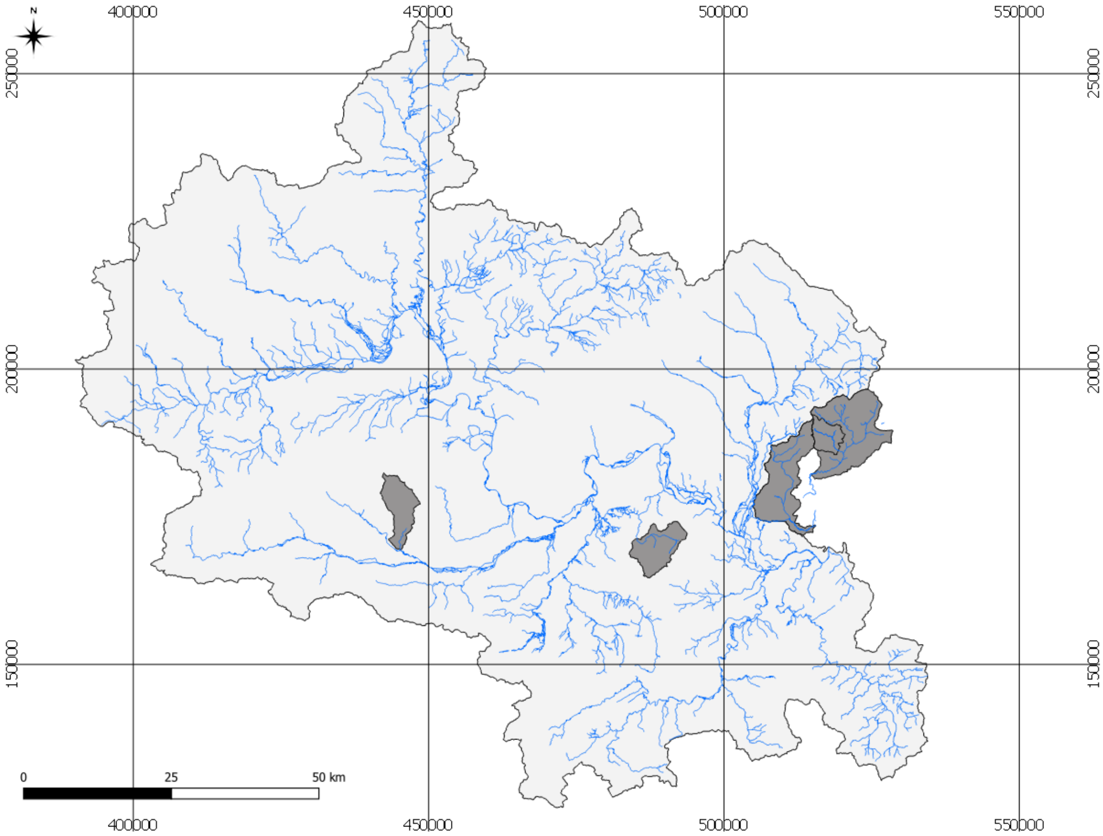


Figure 3-10: Map of Thames river basin in white and five catchments in grey selected for analysis in white.

All five of the catchments are shown on Figure 3-10 in grey. Three heavily urbanised catchments were selected (NRFA: 39094, 39096, 39131), one catchment was selected with very little urbanisation (NRFA: 39033) and the final catchment was selected as it is a built up town and will be used for more analysis in Chapter

5 (Bracknell, NRFA: 39052). The three largely urbanised catchments were then selected based on area in order to obtain a variety of catchments, ranging from a smaller catchment 21.8 km^2 to a larger catchment 146 km^2.

3.3.1 Hydro-meteorological data

The hydro-meteorological daily observed data consisted of: (i) catchment average precipitation i, (ii) average river flow (q_{obs}) and (iii) potential evaporation data (E_p). The daily observed precipitation data were obtained from the CEH-GEAR data set spanning a 50 year period (1961-2012) Keller et al. (2015). The daily potential evaporation data was obtained from the Climate, Hydrology and Ecology research Support System (CHESS) spanning a 50 year period (1961-2012). The daily river flow data were acquired from the National River Flow Archive (NRFA) spanning a 38 year period (1975-2013). The daily rainfall and evaporation data had no missing observations. However three river flow data sets (39094, 39096, 39131) did have missing observations. But none of the data was missing for the period 1995-2004, this period was subsequently chosen as the test period, as continuous data was needed. Each model structure can calibrated run with missing flow observations, by either removing the missing time steps, or having the objective function ignore the time period. If the data is missing rainfall observations a value of zero can be taken for this period, but it will reduce model accuracy.

The URBEXT$_{2000}$ catchment descriptor Bayliss et al. (2006) was used for this chapter, the subscript 2000 denoting that the 50m $\times$ 50m land-cover data that was used to construct the index refers to the period between the years of 1998-2000. URBEXT$_{2000}$ uses a contribution of both urban and sub-urban land-cover classes, with the urban land-cover consisting of roofs, roads and man-made structures, whereas the sub-urban section is a mix of vegetation and semi-built up areas. For a catchment URBEXT$_{2000}$ is defined as the fraction of the catchment covered in urban land-use, with half of the sub-urban area counted as urban land-cover as it is assumed that half of the sub-urban area is made up of vegetation e.g (gardens or parks) Bayliss et al. (2006).

3.3.2 Model Calibration

All three of the model structures are calibrated using the SCE optimisation algorithm outlined in Chapter 3.2.6. A Klemeš split-sample test was used for this study, splitting the 10 years of data (1995-2004) into a five year calibration period (1995-1999) and a five year validation period (2000-2004). The years 1995-1999 were used to calibrate each model on each catchment in order to obtain a parameter set. This parameter set for each respective catchment and model is then applied to the validation period of years 2000-2004 in order to obtain simulated data (q_{sim}). Performance criteria can then be obtained for each catchment and model by comparing the simulated and observed data .

3.3.3 Performance criteria

Two performance criteria were chosen to assess model performance. The first criteria, the well-known Nash-Sutcliffe efficiency statistic NSE Nash and Sutcliffe (1970) defined as :

$$NSE = 1 - \frac{\sum_{t=1}^{n}(q_{obs} - q_{sim})^2}{\sum_{t=1}^{n}(q_{obs} - \bar{q}_{obs})^2} \tag{3.113}$$

with $\bar{q}_{obs}$ denoting the mean of the observed river flow. Values of NSE lie between one and $-\infty$, with a value of one indicating perfect fit, i.e $q_{obs} = q_{sim}$. The second criterion to be used will be the mean absolute error MAE defined as

$$MAE = \frac{\sum_{t=1}^{n}|q_{obs} - q_{sim}|}{n} \tag{3.114}$$

Whereas for NSE, one is the perfect fit, MAE's perfect fit is a value of zero, with MAE's values ranging from 0 to ∞.

3.3.4 Results

Performance of the models using Nash-Sutcliffe efficiency

Figure 3-11 shows the NSE values obtained for both calibration (a) and validation (b) periods across all five test catchments. The purple triangles are urban Framework 1, the red circles are urban Framework 2 and the black squares are urban

Framework 3. Only one red circle is present due to Framework 2 having large negative values for both calibration and validation periods for all catchments.

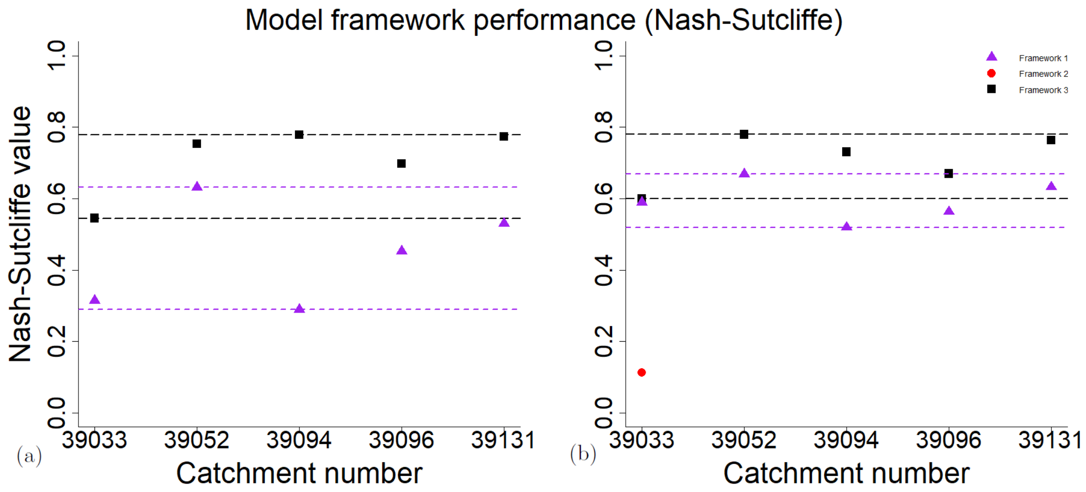


Figure 3-11: Performance of all three model frameworks on the five selected catchments applying the NSE, Framework 1 (purple triangles), Framework 2 (red circle), Framework 3 (black squares).The left hand figure (a) is calibration and right hand figure (b) is validation period.

The results from Figure 3-11 show that for the rural catchment (39033) the performance of Framework 1 and 3 is similar for the validation period. The performance of Framework 1 is both consistent and acceptable for all of the validation periods. The difference between the largest and the smallest NSE was 0.15. The performance of Framework 3 exceeds the performance of Framework 1 for all of the catchments. Calibration for Framework 1 is inconsistent with values ranging from 0.28 (39094) to 0.63 (39052), whereas the validation results were consistent. This indicates that the framework is inconsistent but performs adequately. However the performance of Framework 3 for the calibration periods except for 39033 is consistent, difference between the largest and smallest NSE was 0.08.

Performance of the models using Mean Absolute Error criteria

Next, the model performance as measured by MAE is discussed. Figure 3-12 shows the MAE values obtained for both calibration (left) and validation (right) periods for all five test catchments. Similar to Figure 3-11 the purple triangles are urban Framework 1, the red circles are urban Framework 2 and the black squares are urban Framework 3.

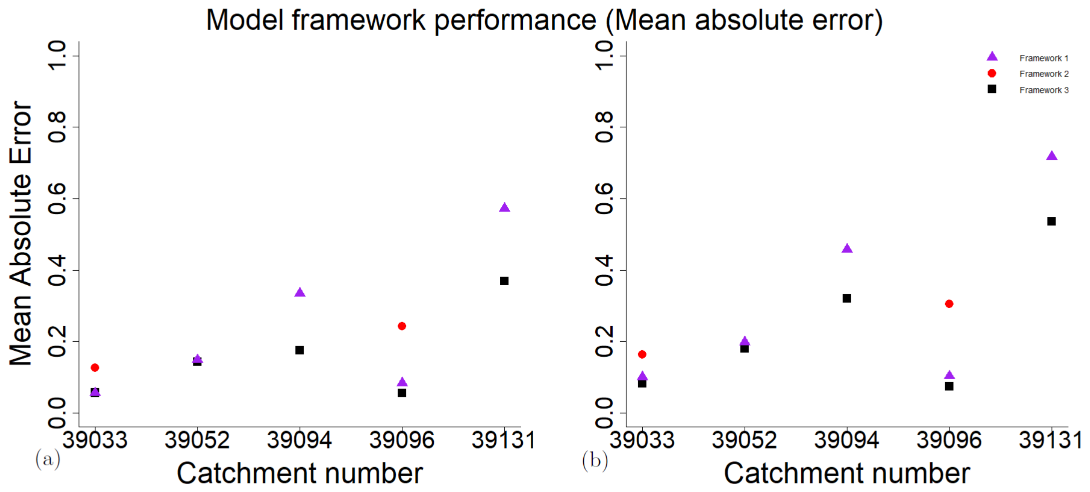


Figure 3-12: Performance of all three model frameworks on the five selected catchments applying the MAE, framework 1 (purple triangles), framework 2 (red circle), framework 3 (black squares). The left hand figure is calibration and right hand figure is validation period.

Since the MAE, is the difference in flow values this criteria still has units unlike the NSE which is dimensionless. As such comparing across catchments can be undertaken with the MAE. Figure 3-12 shows that for three catchments (39033, 39052, 39096) results how that they have very similar calibration and validation performance. For, both Framework 1 and 3 the conclusion that can be drawn is that whilst Framework 3 does perform better than Framework 1 the MAE performance is very similar so a conclusion of which model performs better can not be drawn.

Comparison of simulated and observed data

Figure 3-13 shows the simulated river flow of both Framework 1 and 3 in the catchment 39131 for the validation year 2002. Figure 3-13 shows that both framework 1 and 3 are capable of simulating the base flow. Framework 1 does over estimate during the middle of the year (month April-September), whilst Framework 3 under estimates slightly. During the other days both models are capable of estimating the peaks, with slight under estimation from both models. The simulation results presented and the conclusions drawn are similar to the results in the other two heavily urbanised catchments (39094, 39096).

3.3.5 Conclusion

The results presented in Section 3.3.4 indicate that while urban framework 1 performed acceptable across all catchments, it was not as good as urban framework 3. But Framework 1, assumes that runoff generation of urban areas is independent of soil moisture, however this interaction has been found to be quantitative Redfern et al. (2016). Hence Framework 1 will not be taken forward for further testing in this thesis. Urban Framework 2 was built from the literature assumption that urban runoff is not a flat percentage and scales with the pervious sections of the urban area. Whilst this framework was taking elements from the literature, it's performance was found to be poor, with negative NSE values and very large MAE values. There are two potential reasons for framework 2 performing so poorly. Firstly a software error, such that there is a bug in the code meaning more runoff is generated than the true model values. Secondly, the concept of framework 2 could be wrong. Such that taking 100% runoff at 50% soil moisture generates too much runoff and so bad NSE values are obtained due to squaring these larger values. As such this framework will not be tested further. Finally urban framework 3 performed the best out of all three of the urban frameworks, with consistent performance across all catchments for the NSE. When analysing the hydrographs this framework was able to match the base flow with slight under-estimation in the peaks. As such this urban framework will be selected for future testing.

The next chapter (4) will compare Urban Framework 3 with the single linear reservoir detailed in Section 3.2.5 and the rural model (DAYMOD) on 27 urban catchments in the Thames catchment. The urban model will henceforth be known as the URMOD model. In order to compare the models, two hydrological model comparison tools are developed and analysed. The table below outlines the contribution of the thesis author and the co-author of the coming paper chapter.

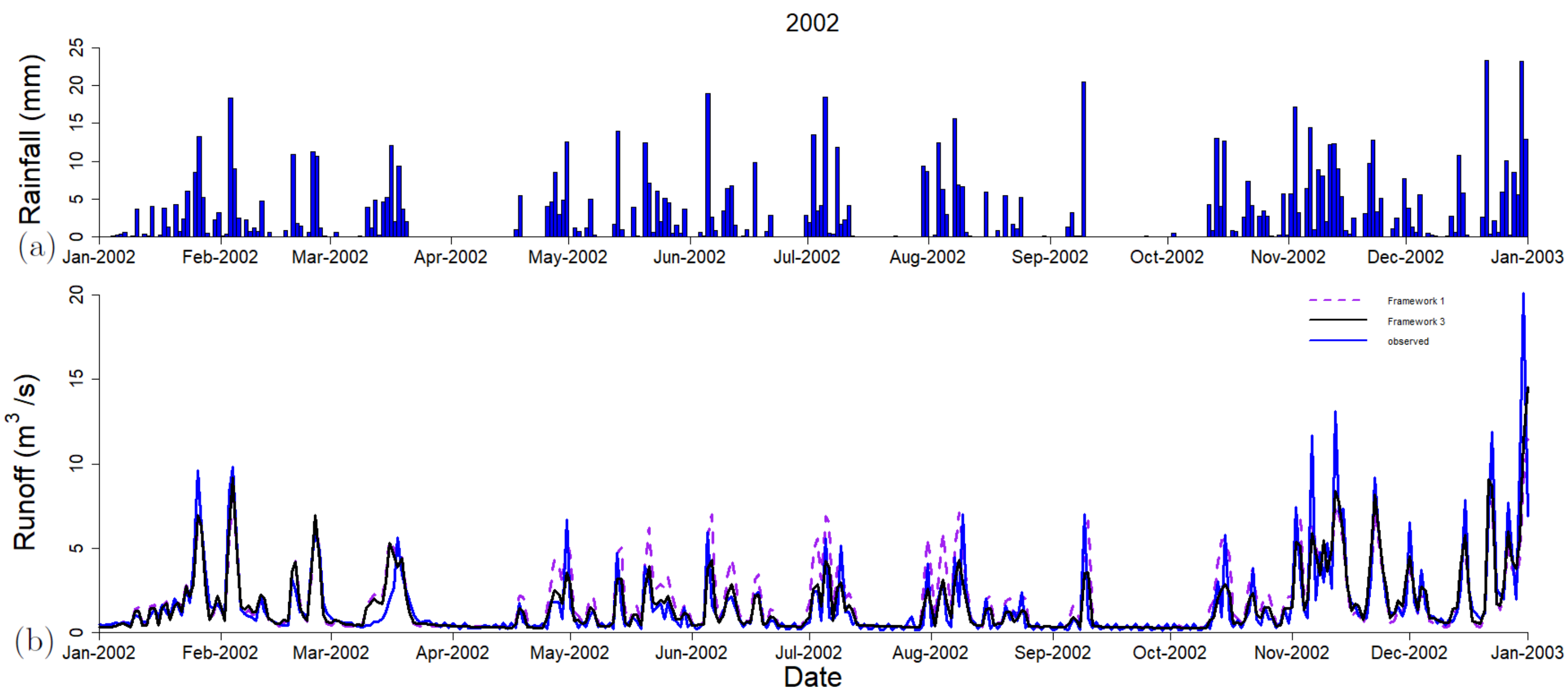


Figure 3-13: Observed and estimated river flow and rainfall. Top figure (a) observed rainfall (mm), bottom figure (b) observed river flow (blus solid line), framework 1 estimated river flow (pruple dashed line), framework 3 estimated river flow (black solid line)

<table>
<tr><td colspan="10">This declaration concerns the article entitled:</td></tr>
<tr><td colspan="10">Operational model comparison techniques for rainfall-runoff models.</td></tr>
<tr><td colspan="10">Publication status</td></tr>
<tr><td>Draft</td><td>✓</td><td>Submitted</td><td></td><td>Review</td><td></td><td>Accepted</td><td></td><td>Published</td><td></td></tr>
<tr><td colspan="2">Publication details (reference)</td><td colspan="8">James Fidal, Thomas Kjeldsen. Draft format with intention to submit to Journal of Hydrologic Engineering</td></tr>
<tr><td colspan="2">Candidate's contribution to the paper (detailed, and also given as a percentage).</td><td colspan="8">The author of this thesis contributed to the entire paper, formulating the ideas, defining the methodology used, application of the models and preparation/ writing of the manuscript. Data collection was not handled as part of this paper, relevant data collectors acknowledged in the paper. Each author's exact contribution is detailed below

Formulation of ideas: 80% James, 20% Thomas
Model application: 100% James
Analyse of model output: 90% James, 10% Thomas
Preparation of manuscript 80% James 20% Thomas</td></tr>
<tr><td colspan="2">Statement from Candidate</td><td colspan="8">This paper reports on original research I conducted during the period of my Higher Degree by Research Candidature.</td></tr>
<tr><td colspan="2">Signed</td><td colspan="2">J.Fidal</td><td colspan="2">Date</td><td colspan="4">October 3 2018</td></tr>
</table>

Table 3.4: Statement of authorship paper:
Operational model comparison techniques for rainfall-runoff models.

Chapter 4

Operational model comparison techniques for rainfall-runoff models.

4.1 Abstract

Comparing the performance of hydrological models is an important and much debated issue in operational hydrology. Multiple methods exist to compare model performance, mainly by focusing on selected performance criteria, which is arguably too simple (Vogel and Sankarasubramanian, 2003). Two new hydrological model comparison tools are proposed in this paper, comparing model performance both for a single catchment and across a group of catchments. The first method is in the form of a significance Z-test using the output of a jackknife split-sample calibration method (Quenouille, 1956) on a single catchment. The second method is a binomial hypothesis test approach considering model performance across a range of catchments. Model performance is assessed for two nested models across a range of urban catchments in the river Thames basin.

4.2 Introduction

Hydrological modelling plays a key role in water management, with many practical applications such as extending stream flow records, predicting future river flows, and simulating river flows in ungauged catchments e.g (Beven, 2011). Deciding if a particular model can adequately represent observed data is typically based on a performance criterion such as the Nash-Sutcliffe efficiency (NSE), Root Mean Square Error (RMSE) or the coefficient of determination (R^2) to compare the observed and simulated runoff. However, Legates and McCabe (1999) argued that simply applying and presenting these criteria is too simple and potentially misleading basis for model selection. They compared a number of performance criteria and concluded that performance criteria can be misleading due to sensitivity to extremal values and insensitivity to additive and proportional differences between model simulation and observed data. They concluded that performance criteria should be used in conjunction with other methods to evaluate model performance. Issues arising when comparing model performance using performance criteria where also explored by Schaefli and Gupta (2007), who concluded that simply relying on the Nash-Sutcliffe efficiency alone is not sufficient to validate a model, as difference in performance at low flows and peaks are lumped together and the performance can not be captured within a singular value.

Similarly Krause et al. (2005) argue that no single performance criterion can be considered singularly 'best' as advantages and disadvantages are evident for all criteria, for example the Nash-Sutcliffe efficiency and the coefficient of determination being sensitive to peak flows. Weglarczyk (1998) highlighted the dangers of using multiple performance criteria due to the interdependence between them. Similar to these studies Vogel and Sankarasubramanian (2003), Pappenberger and Beven (2006), Mishra (2009) and Pechlivanidis et al. (2010) all conclude that selecting one single performance criterion, or in some cases using several performance criteria, to determine model performance can be misleading, and that alternative methods such as: hydrograph analysis, covariance validation procedure or uncertainty analysis techniques such as stepwise regression analysis and entropy analysis should be considered. Studies have attempted to incorporate the aforementioned methods into model assessment studies, such as Bouffard (2014), who compared two models and used both performance criteria and uncertainty analysis. Anh et al. (2010) compared three models using three different performance criteria and graphical analysis on a single catchment. Similarly, Refsgaard and Knudsen (1996) used a combination of performance criteria and plotting observed and model simulated data to compare three models of varying complexity on three different catchments. Whilst all of these studies have applied various techniques to strengthen the scientific basis for model selection, they all rely on performance criteria. Therefore, new methods need to be developed in order to expand current applications of performance criteria to address short-comings yet remain operationally useful. One such method is presented by Ewen (2011) in the form of a dynamic programming algorithm, the Hydrograph Matching Algorithm, which create a visual performance criterion. The Algorithm is applied to selected observed and simulated data presented in the form of hydrographs in order to measure differences in amplitude and timing errors. This method is described by the authors as an extension to performance criteria. They also concluded that performance criteria is a powerful tool in analysis but needs not simply be applied but interpreted such that the choice of which performance criteria are being used and why. The use of performance criteria can be a misleading tool, but it is a tool that is widely used and accepted in hydrological research and operational hydrology. However, a strengthening of model comparisons is needed to be able to apply performance criteria.

Another approach to model performance evaluation is using Hydrological signatures. Hydrological signatures are index values, derived from simulated and observed hydrological data. They are designed to represent relevant information about hydrological behavior. They were first introduced in Gupta et al. (2008) who argued that a diagnostic style approach to model evaluation should be performed over simply viewing a hydrograph or applying performance criteria. Many advancements to hydrological signatures has been conducted with McMillan et al. (2016) outlining guidelines to selecting hydrological signatures and example signatures. With hydrological signatures now being using in model calibration (Euser et al., 2013), model selection ((Clark et al., 2011), (McMillan et al., 2011)), and classification of catchments (Wagener et al., 2007).

Klemeš (1986) proposed a method to compare models by splitting the period of available data into periods one calibration and one validation, called the split-sample test. This method is a standard now in most hydrological model applications (Andréassian et al., 2009), evidenced from studies such as Refsgaard (1997), Ewen (2011), Donnelly-Makowecki and Moore (1999). Ewen and O'Donnell (2012) explored the idea of split-sample calibration and validation further by splitting the calibration period into two sections instead of one: a first calibration period, a second calibration period, and finally a validation period. They concluded that whilst this methodology did improve model simulation results, further research was still needed to determine if the methodology will work on different catchments and different types of storms.

Gharari et al. (2013) presented an alternative method to a singular calibration period by suggesting splitting the calibration period into a number of sub-periods and generating multiple parameter sets. The performance of each parameter set is then calculated and compared in order to determine the single best parameter set. However this methodology only generates multiple calibration periods not validation periods. An advancement to this calibration methodology is presented in this paper based on a jackknife style methodology. The jackknife is defined as a systematic resampling technique as opposed to the bootstrap which is random resampling. Building upon (i) the ideas used to determine parameter variability introduced by, Jones and Kay (2007) and Selle and Hannah (2010), and (ii) splitting the calibration period into subsets Gharari et al. (2013) this paper will introduce a new jackknife methodology to calibrate and validate two different models in order to develop a more robust calibration/validation methodology.

Whilst this study will not directly introduce a new performance criterion, it will instead present two easy-to-use methodologies to determine the difference in performance between hydrological models. These tools will be reliant on existing and widely used performance criteria. The first method will be a performance measure formulated as a paired Z-test for analysing individual catchment performance between two models. The second method is a binomial approach to test for statistical significant between model performance across a group of catchments.

4.3 Model comparison techniques

4.3.1 Calibration and Validation

The first step needed to determine model performance is to define a calibration and validation period. The calibration period is defined as the span of observed data used for calibration of model parameters. The validation period is defined as the period of data in which a comparison of observed and simulated data is undertaken to determine if the model is capable of making accurate simulation when applied outside of the calibration period. Calibration and validation periods which do not overlap are used, as a model needs to be validated on data independent of calibration data in order to show that it has the capacity to predict data and not simply mimic calibration data.

4.3.2 A Jackknife calibration method

The jackknife re-sampling technique developed by Quenouille (1956) is a systematic sampling method, which was originally designed for exploring bias estimation but can also provide estimates of variance which is the primary reason for using the method in this study. The jackknife method used here is an adaptation of the approach used by Jones and Kay (2007), to quantify model parameter uncertainty, but is used here as part of the calibration and validation to assess uncertainty of the performance criteria. The jackknife methodology will be employed in this paper in order to generate multiple performance criteria for two different rainfall-runoff models applied to a single catchment. A paired Z-test method will then be applied to these multiple performance criteria in order to determine if a significant difference in model performance between the two models can be detached for a given catchment. This process will be repeated for multiple catchments, and a second a binomial hypothesis test method will be applied to test significance between the difference in performance of two models across multiple catchments. Whilst this methodology can be applied to more than two models, by comparing each with each other, this paper will be focused on presenting the methodology.

Next, the method to generate the multiple performance criteria will be presented. Denote M_1 as a model that requires parameter calibration, using a set of observed hydrological data (runoff, rainfall, potential evaporation) of length N. The jackknife calibration is based on the following process: first split the calibration data set into $j = 1, ..., n$ equal length (N/m) non-overlapping periods. The first sub-period has the same first value as the full set and the final sub-period has the same final value as the full set. Next, calibrate M_1 using the data in the first sub-period $(j = 1)$, resulting in a set of model parameters θ_1. Next use M_1 with parameter set θ_1 to simulate runoff on the remaining $n - 1$ sub periods. A performance criterion Z_1 can be obtained for the validation period by comparing the model simulated data and observed data in the validation period $j = 2, ..., n$. Next, the model is calibrated on the second sub-period, $j = 2$, and validated on the remaining $j = 1, 3, ..., n$ sub-periods resulting in a new value of performance criterion, Z_2. This process is then repeated systematically until the models have been calibrated on all individual sub-periods, each time using a different sub-period for model calibration and validation. For each iteration, a performance criterion Z_j and parameter set θ_j is obtained, such that a set of $Z_j, j = 1, ..., n$ performance criteria and $\theta_j, j = 1, ..., n$ datasets are obtained. Figure 4-1 shows an example of the jackknife process for a data set of length 30 years, a calibration length of 2 years, a validation length of 28 years, hence $Z_j, j = 1, ..., 15$ performance criteria would be obtained.

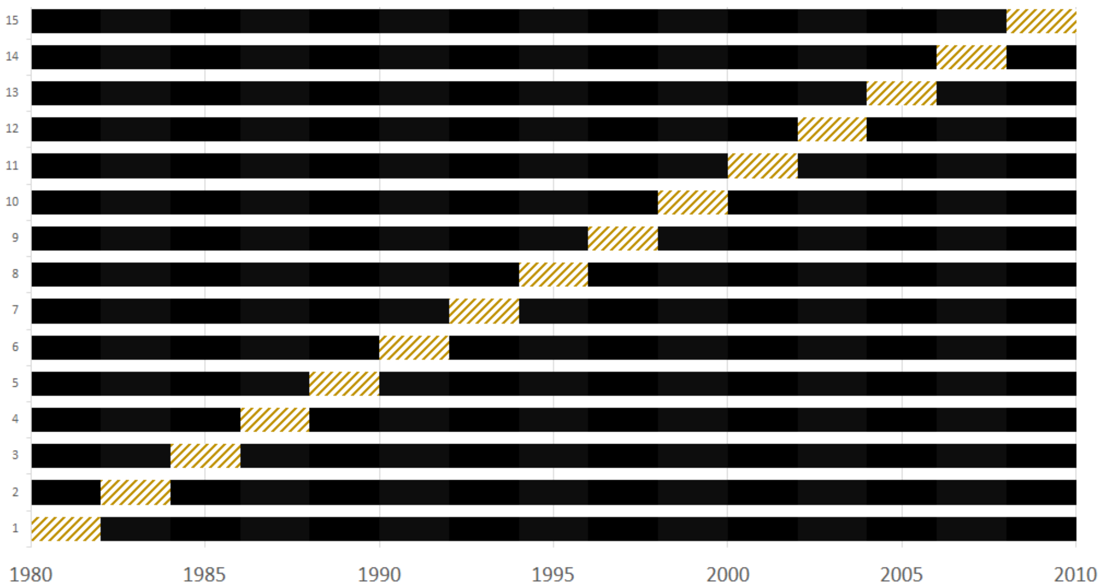


Figure 4-1: Example Jackknife method for data sets of length $N = 30$, $m = 2$, $v = 28$, $j = 15$. Gold blocks represent the calibration period and black represents validation period.

The set of performance criteria $Z_j, j = 1, ..., n$ can be used to assess the uncertainty in model performance. The next two sections will outline how the performance criteria sets obtained can be used to evaluate model performance.

4.3.3 Paired Z-test method

The first method to evaluate model performance will be based on the concept of a paired Z-test. This method is used to compare two performance criteria sets obtained by applying two different models M_1 and M_2 to the same data set through the jackknife calibration method. Applying the jackknife methodology from section 4.3.2 with two models M_1 and M_2 to obtain two sets of performance criteria $Z_{1,j}$ and $Z_{2,j}$ such that both sets are of equal length with the subscript denoting which model it is obtained from. The difference of each observation of the sets is given as $Z_{d,j} = Z_{1,j} - Z_{2,j}, j = 1, ..., n$. The mean of Z_d can be obtained

$$\overline{Z_d} = n^{-1} \sum_{j=1}^{n} Z_{d,j}. \tag{4.1}$$

The associated variance estimate as stated by Efron (1982) can also be calculated as:

$$Var\{\overline{Z_d}\} = \frac{n-1}{n} \sum_{i=1}^{n} (Z_{d,j} - \overline{Z_d})^2. \tag{4.2}$$

The mean value is assumed normal distributed, and consequently the 95% confidence interval of the mean performance measure $\overline{Z_d}$ of each model defined as:

$$\left(\overline{Z_d} - 1.96\sqrt{Var\{\overline{Z_d}\}}, \overline{Z_d} + 1.96\sqrt{Var\{\overline{Z_d}\}} \right). \tag{4.3}$$

A hypothesis test can be formed such that the null hypothesis (H_0) states that the performance of models M_1 and M_2 is the same, i.e.:

$$H_0 : \overline{Z_d} = 0.$$

whereas the alternative is defined as model performance being different, i.e one model is performing better than the other,

$$H_1 : \overline{Z_d} \neq 0.$$

In order to determine which hypothesis to accept and reject, the confidence interval for $\overline{Z_d}$ is interpreted such that if the interval contains zero then the null hypothesis of equal performance can be accepted, whereas if the interval does not contain zero then the null hypothesis can be rejected, thus indicating a significant difference in model performance.

4.3.4 Binomial hypothesis test

Whilst Section 4.3.3 outlines a method to compare models on a singular catchment, the binomial hypothesis test can be used to compare model performance across a group of catchments. The test will utilise a success/ failure approach to compare two models; either a model outperforms the other or it does not. Whilst it has not been reported in the literature of models performance being equal, model performance can be similar or very close to equal. However since the nature of the Binomial distribution is success/ fail the case where both model performance criteria are exactly equal is not taken into account, as this case is considered an unlikely occurrence. In order to apply a binomial hypothesis test, independence between model performance assessments in each sub-period has to be assumed, meaning each calibration of a model must be independent and holds no prior calibration such that parameter calibration is not influence via previous calibrations.

The methodology described in Section 4.3.2 a common record length is assumed, however as discussed in Section 4.6 a varying record for each catchment and varying sub-periods for each catchment can be used. Applying the jackknife calibration/validation method to two models M_1 and M_2 on each catchment in turn in a group of c catchments will result in c sets of performance criteria each containing n_j elements $Z_{1,j,l}$, $j = 1, ...n$, $l = 1, ..., c$ and $Z_{2,j,l}$, $j = 1, ...n$,$l = 1, ..., c$, with subscript l denoting the catchment number. For each catchment the difference between model performance is calculated for each of the $j = 1, ..., n$ validation periods, $Z_{d,j,l}$ can be obtained so $Z_{d,j}$ performance criteria are obtained. The respective means of each of the difference sets can be obtained for each catchment $\overline{Z_d}$, resulting in c performance criteria means. This process is displayed in Figure 4-2

A hypothesis test can be formed such that the null hypothesis (H_0) assumes that there is no difference in model performance across the c catchments, and therefore the probability (p) that either model outperforms the other is assumed

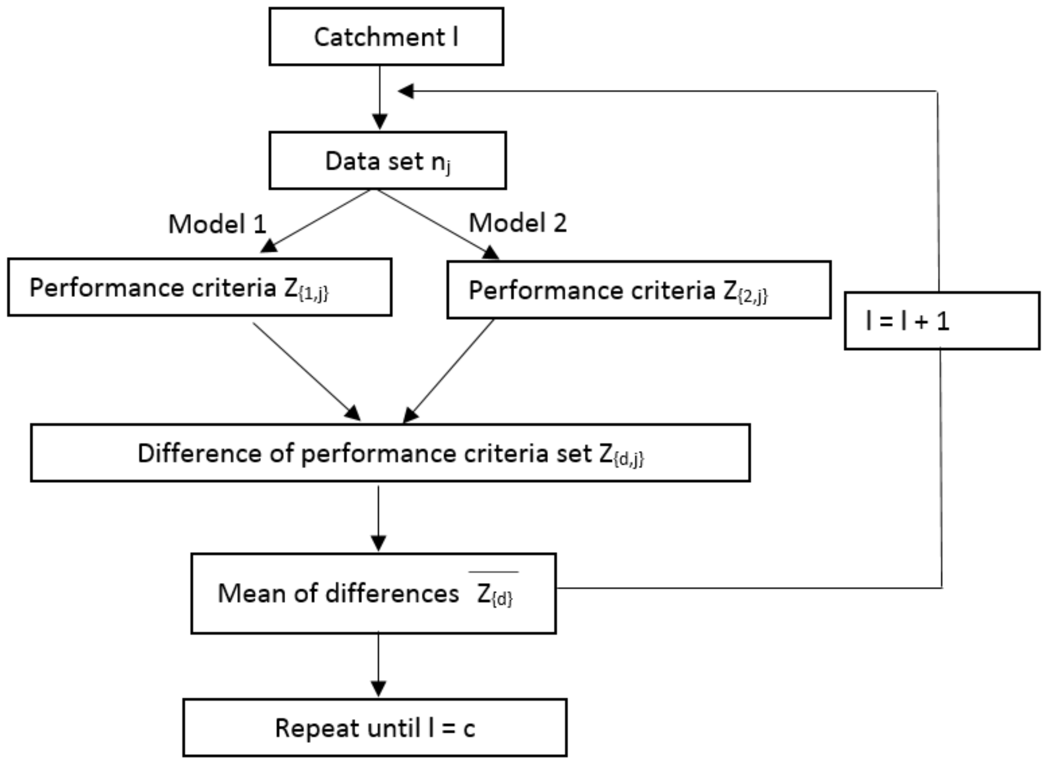


Figure 4-2: Data process for l catchments and two models.

to be half, i.e.:

$$H_0 : p = 0.5. \tag{4.4}$$

The alternative hypothesis (H_1) can be set up as either a two tailed test, if a test for either model outperforming the other, or a one tailed test, if a test for a specific model outperforming the other is preferred. The case presented here is the two tailed test of model performance being different. The alternative hypothesis is given as :

$$H_1 : p \neq 0.5. \tag{4.5}$$

Let c denote the number of trails, such that a trail is defined as the mean difference of performance criteria for two models $\overline{Z_d}$ on a particular catchment. Define a success as M_1 outperforming M_2 ($\overline{Z_d} > 0$), whereas a failure is M_2 outperforming M_1 ($\overline{Z_d} < 0$). Let V be a random variable defined as the number of catchments where M_1 outperforms M_2 (successes). Thus the probability of v

instances where M_1 outperforms M_2 is a binomial distribution $B(c, p)$ and given as;

$$P\{V = v\} = \binom{c}{v} p^v (1-p)^{(c-v)}. \tag{4.6}$$

A hypothesis test can be formed for a predefined significance level (α) e.g. $\alpha = 5\%$. The observed number of success v is compared to the critical interval defined as $v \leq B(c,p)_{\frac{\alpha}{2}}$ and $v > B(c,p)_{1-\frac{\alpha}{2}}$ where subscript $(\alpha/2)$ and $(1-\alpha/2)$ signify quantiles of the binomial distribution. The values obtained from Equation 4.6 will be denoted p-values such that if the p-value falls within the critical interval then the null hypothesis can be rejected such that there is a difference in model performance, whereas if the p-value does not fall within the critical interval then the null hypothesis can be accepted such that there is no difference in model performance. Figure 4-3 is an example of the method for a two-tailed example let $c = 27$ and $p = 0.5$, then to achieve a significant difference in model performance at $\alpha = 5\%$ significance level, M_1 would have to outperform M_2 in 19 out 27 catchments or M_2 would have to outperform M_1 in 8 out of 27 catchments as indicated by the red section of Figure 4-3.

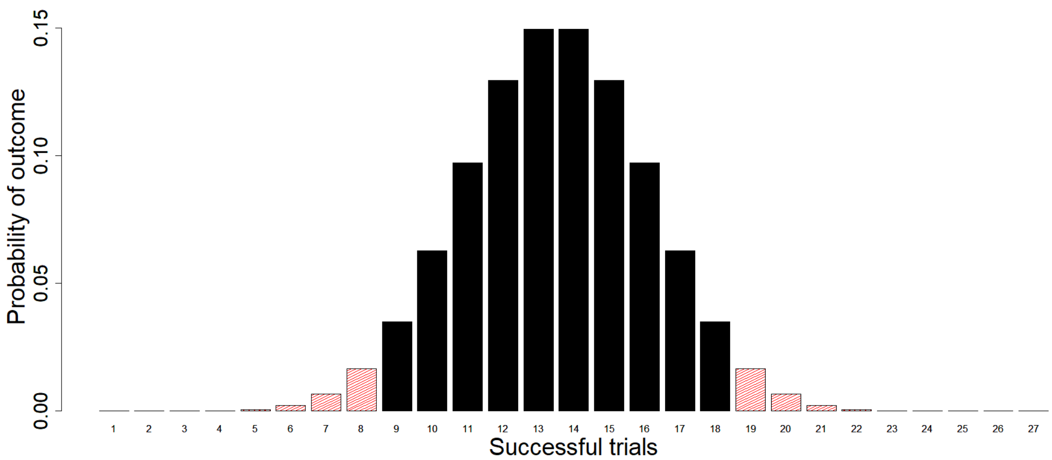


Figure 4-3: Binomial distribution $c = 27$ and $p = 0.5$, the red areas indicate significance.

An example case study is presented in the next section (Section 4.4) to test the methods, with the final section (Section 4.6) discussing the advantages and disadvantages associated with these methods.

4.4 Case study: The Thames Catchment

The two hydrological model comparison tools developed in the previous sections were tested using two conceptual rainfall-runoff models on a set of $c = 27$ gauged catchments located within the Thames catchment.

4.4.1 Model description

The two models used for this case study are URMOD (M_2) and DAYMOD (M_1) (Kjeldsen et al., 2005). URMOD is an extension of DAYMOD containing an additional set of model components to account for urban land-use, resulting in a nested model structure where M_2 is a simpler version of M_1. Both models are a lumped conceptual parameter-parsimonious rainfall-runoff model (As explained in Chapter 3) with URMOD having eight calibrated parameters (since it is using the single reservoir not the parallel linear reservoir which adds an additional parameter) whereas DAYMOD has seven.

The models represent two main processes (i) infiltration and runoff and (ii) channel routing. The infiltration and runoff generation is based on a conceptual soil column approach, such that the precipitation that does not infiltrate is turned into runoff. The runoff generation in DAYMOD is dependent on the soil moisture in the conceptual soil column such that as the column fills more runoff is generated. The runoff generation in URMOD is split into two contributions, one from the rural ares and the second from the urban areas. Runoff generation from the rural areas is the same as DAYMOD, whereas urban runoff generation is determined via a calibrated scaling term, with a larger scaling term resulting in more urban runoff being generated. The second process within the models is the channel routing. This is based on single linear reservoir such that routing for the rural areas is achieved via a linear reservoir concept with a proportion of the runoff routed through a baseflow reservoir then routed through a surface flow reservoir. In contrast, the proportion of runoff designated as surface flow is just routed through the surface flow reservoir. The baseflow and surface flow is then combined to be the rural runoff at the catchment outlet. The urban routing within URMOD routs the urban runoff through a separate surface flow reservoir, and then combines with the rural runoff to become total runoff at the catchment outlet. The two models require observed rainfall, runoff and potential evaporation, in order to calibrate the 8 parameters (DAYMOD has 7 calibrated parameters). Both models are calibrated using a shuffled evolution complex algorithm (Duan et al., 1993).

4.4.2 Catchment Selection

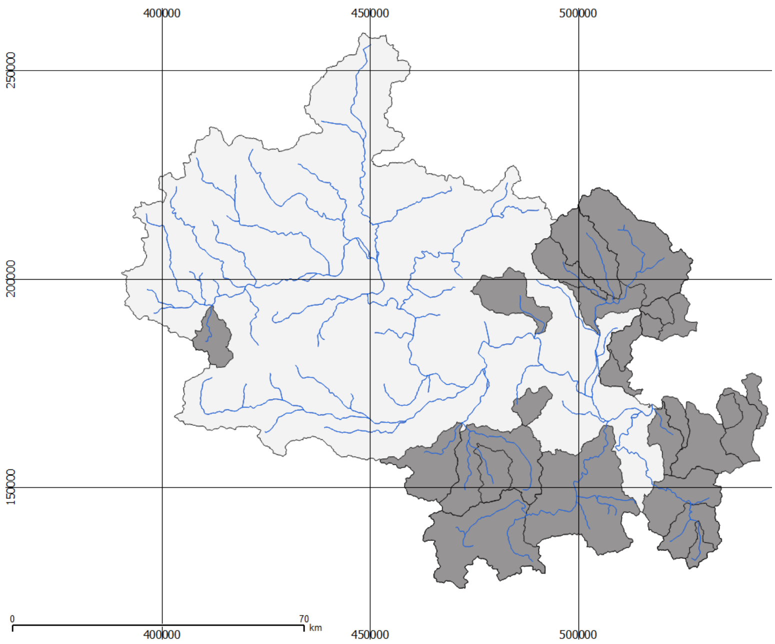


Figure 4-4: Thames river catchments, entire Thames basin in white and selected catchments in grey.

An initial set of 112 catchments were assembled from within the Thames catchment for which long-term daily rainfall and runoff data are available from the National River Flow Archive (NRFA). This initial set was reduced to a subset of 27 catchments based on the condition that the fraction of urban land cover had to be larger than 5% of the catchment to ensure a meaningful comparison of URMOD and DAYMOD. Futhermore, each catchment needed good quality data for a 30 year period 01/01/1980 to 31/12/2009. The resulting 27 catchments ranged in size from 21.8 km^2 to 904 km^2 with fractional urban land cover values ranging from 5.34% to 54.75%. In Figure 4-4 the 27 catchments are highlighted in grey.

The hydro-meteorological data used in this study consist of: catchment average daily precipitation (i), average daily river flow (q_{obs}), and daily potential evaporation data (E_p). Runoff data at a daily time step were acquired from the National River Flow Archive (NRFA) spanning 30 years from 01/01/1980 up to 31/12/2009. The precipitation data were obtained from the CEH-GEAR data set (Keller et al., 2015) covering the same 30-year period. Finally, evaporation data was obtained from the Climate, Hydrology and Ecology research support system (CHESS) (Robinson, 2016). The runoff data was quality controlled by removing the missing data rather than estimating values, checks were made to ensure that there were no major gaps of multiple weeks worth of data missing. When continous weeks of data is missing this is defined as a major gap of missing data.

One important criterion for the urban model is determining the percentage of urban land-use in a catchment. For this study the URBEXT_{2000} catchment descriptor (Bayliss et al., 2006) was used, where the subscript 2000 denotes that the 50m $\times$ 50m land-cover data used to construct the index refers to land-use data from the period between the years of 1998-2000. URBEXT_{2000} uses a contribution of both urban and sub-urban land-cover classes, with the urban land-cover consisting of roofs, roads and man-made structures, whereas the sub-urban section is a mix of vegetation and semi-built up areas, only half of the sub-urban section is contributed to URBEXT_{2000} as it is assumed that half of the sub-urban section is made up of vegetation such as gardens or parks (Bayliss et al., 2006).

4.4.3 Performance criteria

In order to apply the two methods presented in Section 4.3, a set of performance criteria needs to be selected. Two different performance criteria were selected for this study: (i) the Nash-Sutcliffe efficiency statistic (NSE) (Nash and Sutcliffe, 1970), and (ii) the Mean Absolute Error (MAE). Consider a time series of observed runoff $q_{obs}(t), t = 1, ..., n$, and the accompanying simulated runoff $q_{sim}(t), t = 1, ..., n$ obtained from a calibrated rainfall-runoff model. Definitions and reasons for choosing each criteria is presented below.

The Nash-Sutcliffe efficiency statistic (NSE) statistic is defined as

$$NSE = 1 - \frac{\sum_{t=1}^{n}(q_{obs}(t) - q_{sim}(t))^2}{\sum_{t=1}^{n}(q_{obs}(t) - \bar{q}_{obs}(t))^2} \tag{4.7}$$

The range of possible NSE values spans $-\infty$ and one, with a value of one indicating perfect fit, i.e $q_{sim} = q_{obs}$ for all nj observations. This criterion was selected because its widespread use in hydrology. An often cited problem with this criteria is that it is sensitive to extremal events due to it squaring differences as discussed by Krause et al. (2005) and Legates and McCabe (1999). Because of this potential problem an absolute value criteria shall be chosen alongside the NSE.

The second criteria adapted in this study is the Mean absolute error MAE defined as

$$MAE = \frac{\sum_{t=1}^{n} |q_{obs} - q_{sim}|}{n} \tag{4.8}$$

Possible values of MAE range from 0 to ∞, where a perfect fit will result in a value of zero. This criteria was chosen because, unlike the NSE, it takes the absolute difference rather than the squared difference and is therefore less affected by high deviations (Willmott and Matsuura, 2005).

4.4.4 Experiment setup

The two methods developed in Section 4.3 were applied in an attempt to compare the performance of the two rainfall-runoff models, described and developed within Section 3.1 and 3.2, URMOD (M_2) and DAYMOD (M_1) when applied to 30 years of observed hydrological data on 27 catchments in the Thames basin. Both models use the same calibration scheme explained in Section 3.2.6 . Two calibration lengths were chosen, 1 year and 2 year, resulting in $n = 30$ and $n = 15$ performance criteria and parameter sets, respectively, for each catchment. These time lengths were chosen in order to obtain suitably large amounts of performance criteria in order to apply the Z-test method. Applying the jackknife calibration method outlined in Section 4.3.2 both the NSE and MAE performance criteria were used to evaluate model performance. The paired Z-test method outlined in Section 4.3.3 is applied to each of the 27 catchments in turn to determine the individual performance of each catchment between the two models. Next the binomial method outlined in Section 4.3.4 was applied to determine the collective performance across the $m = 27$ urban catchments.

4.5 Results

4.5.1 Assessing performance of individual catchments via paired Z-test

This section will explore the difference in performance of the two models at the individual catchment level, using the paired Z-test. Whilst the tools presented in Section 4.3 detailed a method for the validation period, they can also be applied to the calibration period. This is achieved when the optimal model parameters are calibrated, they are applied to the model in order to generate simulated data for the calibration period. This data can then be used to obtain a performance criteria, with the process being repeated for each jackknife calibration period in order to obtain multiple performance criteria. The paired Z-test confidence intervals can then be calculated. The results will be presented such a positive difference in performance criteria ($\overline{Z_d} > 0$) indicates model M_2 (URMOD) performed better than model M_1 (DAYMOD). In contrast negative difference ($\overline{Z_d} < 0$) indicates that model M_1 (DAYMOD) performed better than model M_2. The 95% confidence intervals of $\overline{Z_d}$ are calculated for each catchment using Eq 4.3 and then subsequently plotted. If the confidence intervals cross zero, then the null hypothesis of equal M_1 and M_2 performance can be accepted, whereas if the interval does not cross zero the null hypothesis is rejected such that either M_1 or M_2 has performed better.

1-year calibration

Figures 4-5 and 4-6 show the results of difference in performance of M_2 and M_1 when calibrated on 1-year, and the performance assessed on the 29-year validation periods. The left hand figures (a) show the difference in performance criteria of the calibration period and the right hand side (b) is the difference in performance criteria of the validation period. The black circles indicate that $\overline{Z_d} > 0$. Whilst the red triangles are $\overline{Z_d} < 0$. Figure 4-5 shows the difference in performance when the MAE criteria is applied, whereas Figure 4-6 is the Nash-Sutcliffe performance criteria. The black lines indicate the 95% confidence intervals.

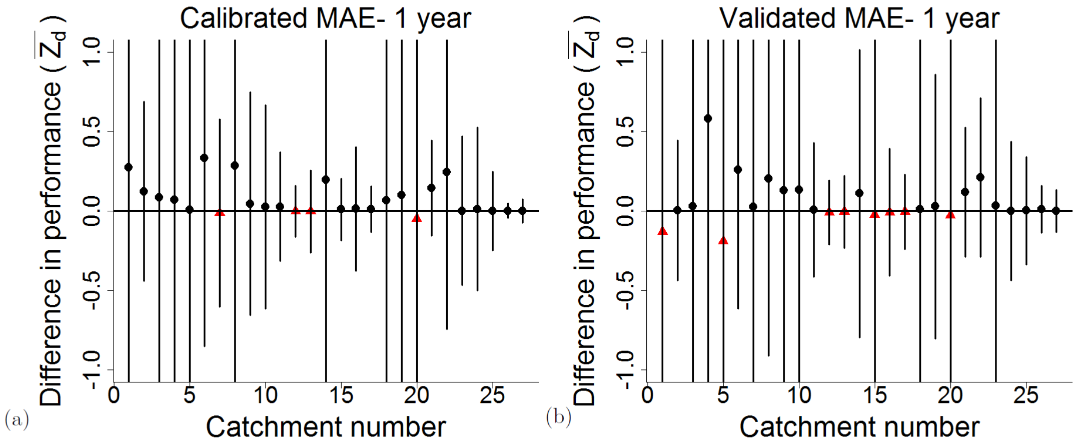


Figure 4-5: Difference in performance $\overline{Z_d}$ for calibration (left hand figure) and validation periods (right hand figure). A calibration period of 1-year was used and MAE performance criteria. Black circles show that $\overline{Z_d} > 0$, red triangles show $\overline{Z_d} < 0$

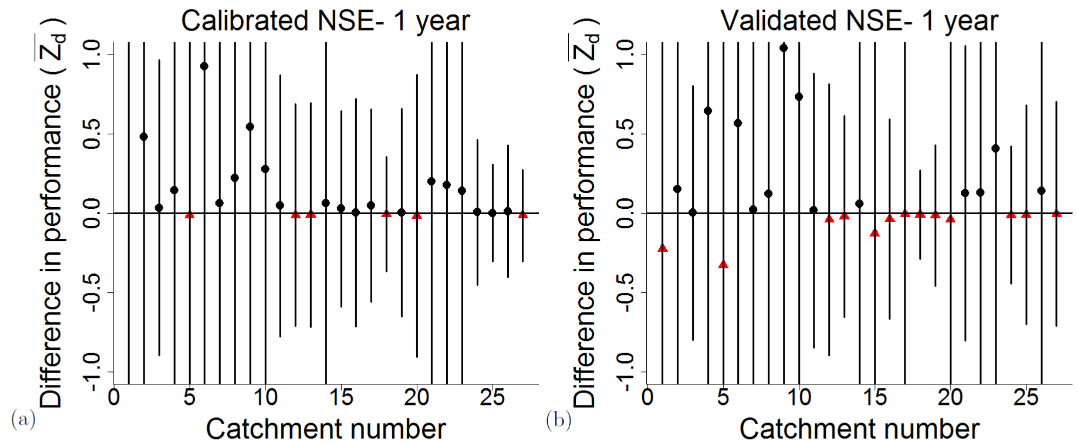


Figure 4-6: Difference in performance $\overline{Z_d}$ for calibration (left hand figure) and validation (right hand figure). A calibration period of 1-year was used and NSE performance criteria. Black circles show that $\overline{Z_d} > 0$, red triangles show $\overline{Z_d} < 0$

The results in Figure 4-5 show that for the MAE calibration period results M_2 outperformed M_1 on 23 catchments. Whilst for the NSE M_2 outperformed on 21 catchments as indicated via Figure 4-6. All of the confidence intervals for the calibration periods include zero indicating that the null hypothesis of equal performance can not be rejected. For the validation period the number of instances where M_1 outperformed M_2 increases. For the MAE criteria (Figure 4-5) it increases from 4 to 8 whereas using the NSE increases the number from 6 to 13. Similar to the calibration period all of the confidence intervals include zero, indicating that the null hypothesis of equal performance can be accepted.

2-year calibration

Figures 4-7 and 4-8 show the results of difference in performance of M_2 and M_1 calibrated on 2-years, with validation period of 28-years. Similar to the results presented in Figures 4-5 and 4-6, the black circles indicate that $\overline{Z_d} > 0$, whilst the red triangles represent $\overline{Z_d} < 0$. Figure 4-7 shows the difference in performance when the MAE criteria is applied, whereas Figure 4-8 contains results obtained using the Nash-Sutcliffe performance criteria. The left hand figures (a) are the difference in performance of the calibration period, while the right hand side (b) is the difference in performance criteria of the validation period. The black lines indicate the 95% confidence intervals.

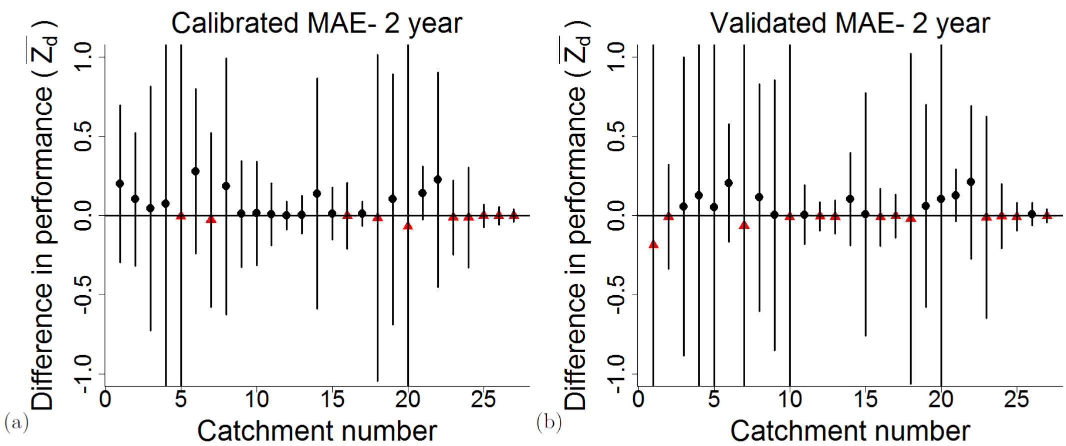


Figure 4-7: Difference in performance $\overline{Z_d}$ for calibration (left hand figure) and validation periods (right hand figure). A calibration period of 2-years was used and MAE performance criteria. Black circles show that $\overline{Z_d} > 0$, red triangles show $\overline{Z_d} < 0$

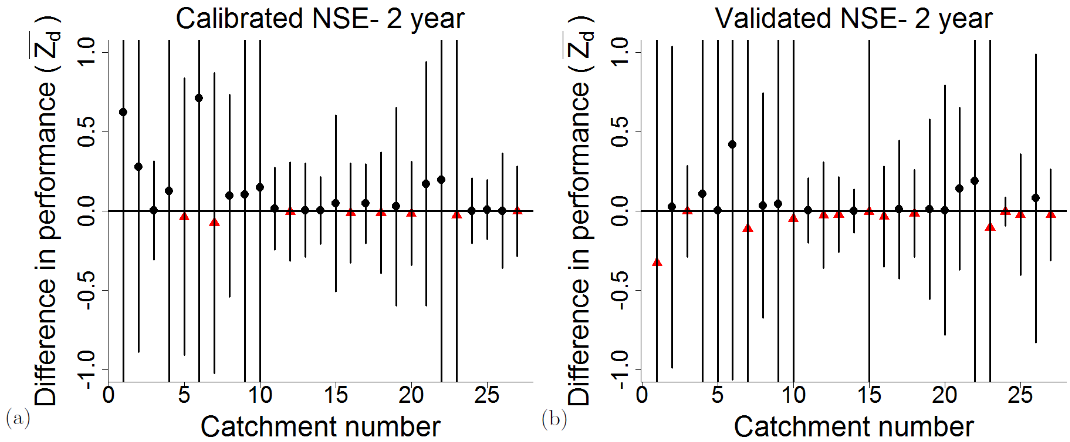


Figure 4-8: Difference in performance $\overline{Z_d}$ for calibration (left hand figure) and validation (right hand figure). A calibration period of 2-year was used and NSE performance criteria. Black circles show that $\overline{Z_d} > 0$, red triangles show $\overline{Z_d} < 0$

Figure 4-7 shows that for the 2-year calibration period model M_2 outperformed M_1 on 17 catchments, whereas for the NSE M_2 the number of catchments where M_2 outperformed M_1, increased to 20. Again all confidence intervals include zero indicating the null hypothesis of equal performance can be accepted. Similar to the 1-year validation, for the 2-year validation the number of catchments of M_1 outperforming M_2 increases, for the NSE an increase from 8 to 13 catchments with MAE increasing from 10 to 13. Similar to the 1 year confidence intervals, all intervals did include zero indicating that the null hypothesis of equal performance can be accepted.

4.5.2 Assessing performance of collective catchments via Binomial hypothesis method

This section explores the difference in performance of the two models collectively across all 27 catchments, using the binomial distribution approach as outlined in Section 4.3.4. The hypothesis test is applied to both the calibration and validation results presented in Section 4.5.1. The hypothesis test is a two-tailed test such that the alternative hypothesis H_1 is that there is a statistically significant difference in the models performance in favour of the model with the majority of outperformances. A p-value approach is taken on a $\alpha = 5\%$ critical interval, such that if the p-value obtained is less than 0.025 the alternative H_1 can be accepted whilst a value larger indicates the null hypothesis H_0 of equal performance can be accepted. Table 6.4 shows the performance when a 1-year calibration is used, whilst Table 4.2 is the performance when a 2-year calibration is used.

Performance criteria	Calibration or Validation	Difference (M_2 outperforming M_1)	p-value	Outcome
NSE	Calibration	21 out of 27 catchments	0.002	H_0 can be rejected such that there is a significant difference in model performance in favor of M_2
NSE	Validation	14 out of 27 catchments	0.149	H_0 of equal performance cannot be rejected
MAE	Calibration	23 out of 27 catchments	0.0001	H_0 is rejected such that there is a significant difference in model performance in favor of M_2
MAE	Validation	19 out of 27 catchments	0.017	the H_0 can be rejected such that there is a difference in model performance in favor of M_2.

Table 4.1: Performance of M_2 and M_1, with out-performances of M_2 over M_1 and the Binomial hypothesis test results for a 1-year calibration period.

The results show that for both 1 and 2 year calibration M_2 performs significantly better in the calibration of the NSE, whereas during the validation the null hypothesis of equal performance cannot be rejected. However the results differ between years when the MAE criterion is used. For both validation periods the null hypothesis again cannot be rejected. When 1-year data is used the H_0 can be rejected such that M_2 performs significantly better during the calibration, but for 2-year the null hypothesis cannot be rejected.

Performance criteria	Calibration or Validation	Difference (M_2 outperforming M_1)	p-value	outcome
NSE	Calibration	19 out of 27 catchments	0.017	H_0 can be rejected such that there is a significant difference in model performance in favor of M_2
NSE	Validation	14 out of 27 catchments	0.149	H_0 of equal performance cannot be rejected.
MAE	Calibration	17 out of 27 catchments	0.063	H_0 of equal performance cannot be rejected.
MAE	Validation	14 out of 27 catchments	0.149	H_0 of equal performance cannot be rejected

Table 4.2: Performance of M_2 and M_1, with out-performances of M_2 over M_1 and the Binomial hypothesis test results for a 2-year calibration period.

4.6 Discussion

The results presented in this study raises a number of issues that need further discussion. Firstly, is the difference in results obtained using MAE and NSE. For the 1-year calibration validation results M_2 had better performing performance criteria in 19 out of 27 catchments when applying the MAE. However when applying the NSE this value was only 14. This further highlights Legates and McCabe (1999) conclusions that the choice of efficiency criteria needs to be made clear. For the NSE 1 year validation results values as low as -9 was obtained. Whilst this could indicate poor model performance on the catchment, it also highlights the erratic nature of the criteria. A method to solve this would be too apply a bounded performance criteria such as the bounded Nash-Sutcliffe. However this low value highlights one problem with the jackknife calibration and validation method, such that a large period of years (29) creates a singular lumped performance criteria. Whilst this singular lumped value is good for testing a models overall performance across a large number of years it cannot show if model performance increases or decreases over the years. If this is preferred then an extension to the presented jackknife calibration and validation method can be employed such that the 28-years could have been split into 14, 2-year periods each with there own respective efficiency criteria. This process is shown in Figure 4-9.

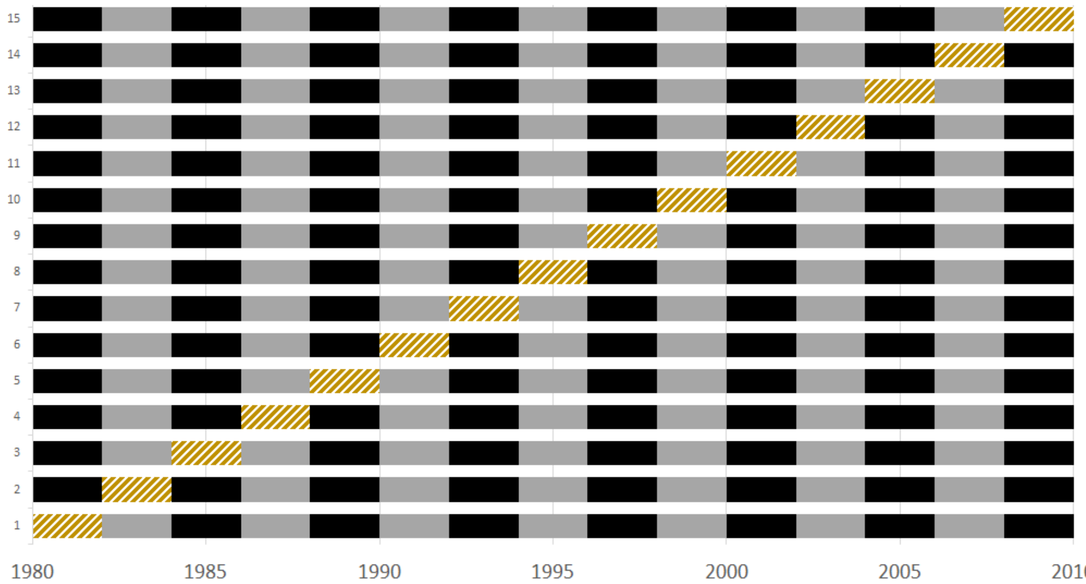


Figure 4-9: Example Jackknife method for data sets of length $N = 30$, $m = 2$, $v = 28$, $j = 15$. Gold blocks represents the calibration period, grey and black represents validation period.

Whilst the jackknife calibration method presented in this chapter is similar to the Klemeš split sample test the jackknife has a number of advantages over the split sample test. Firstly, the jackknife calibration scheme generates multiple calibration and validation results which can be further analysed as opposed to the Klemeš test which only obtains 1 calibration and validation period. However due to the increase in calibration periods the computational time of the jackknife scheme further increases. With multiple calibrations, multiple parameter and performance criteria is obtained which means more rigorous analysis can be conducted using the jackknife method as opposed to Klemeš.

In the application of the jackknife calibration and validation method a common record length was assumed (30 years), with a common sub-period selected (1-year and 2-year). However a varying record-length for each catchment can be used. Varying sub-period length can also be used for different catchments but the length has to be consistent across the entire record length for a singular catchment. However, the total number of sub-periods needs to be considered. As when the the paired Z-test method is applied it as advised by Van Belle (2011) a minimum of 12 data points should ideally be used when estimating variance. One advantage of the jackknife calibration method is that it can generate more parameter sets, which Gharari et al. (2013) argued is preferable. However, this is obtained at the expense of an increase in computational time, due to the need to perform more model calibrations.

As shown in Section 4.5.1, the results indicated that for no catchments were statistically significant indicated by the significance lines not crossing zero, indicating consistent non significance. This could be due to the model's performance being the same, or could indicate that the 28 or 29 years of lumped data creates to much variation due to the inflated variance generated from Equation 4.3. Hence further research is needed in order to see if this method is a viable hydrological comparison tool.

One clear advantage of the proposed Binomial test is the ability to assess if a model is significantly better performing across a large number of catchments. This is a simple-to-use methodology that applies current performance criteria. The binomial approach is also flexible with the jackknife calibration and validation method, rather than comparing multiple catchments the binomial method can be applied to a singular catchment, such that the trials would be denoted as individual years or certain events.

However one issue is raised in the conflicting conclusions between the paired Z-test and the binomial hypothesis test, due to the application of the Z-test indicating no significance between model performance at the individual catchment level, but the Binomial hypothesis test showed that one model performed better across a number of catchments.

This raises the question of whether the binomial hypothesis test can be applied when the catchment performance was shown to be statistically similar. It can because both tests are testing different hypothesis. Since the jackknife Z-test is testing model performance at the individual catchment level, whilst the binomial hypothesis test is testing out performances in a large number of catchments. If in future a method does generate significance for singular catchments then comparing just on significant catchments for the binomial hypothesis test can be applied.

The overall aim of both tests developed within this chapter are to inform decisions within hydrological modelling. The conclusions made about the rainfall runoff models would have differed if the tests developed were not used or if only a single test was used. This is because both tests give differing conclusions on the performance of the models, with the Binomial hypothesis test exploring performance across a number of catchments whereas the jackknife was testing a single catchment. If only one of the tests developed within the chapter are performed a different conclusion could have been drawn. The same could be true if even more different tests are performed.

4.7 Conclusion

This paper presented two easy-to-use techniques for comparing the performance of two rainfall-runoff models, as well as presenting a more robust methodology to calibrate and validate rainfall-runoff models. The purpose of this paper was to show that applying simple statistical methods can add interpretive power when comparing model performance and that model calibration methods can be improved. However whilst these techniques do add a more robust method to test model performance, it is recommenced that these techniques should be used alongside other performance methods such as graphical analysis (hydrographs). One concluding remark on the calibration method is that this method does not improve model structure or calibration algorithms. Moreover this method should be used in the application phase not in the development phase of modelling.

Acknowledgments

The authors would like to thank the National River Flow Archive (NRFA) for providing access to hydrological data. The NERC funded POLLCURB project for providing access to the hydrological and land-use data used in this study (NE/K002317/1). NERC for access to land-use shapefiles, Data owned by NERC Centre for Ecology & Hydrology.

4.8 Bibliography

Andréassian, V., Perrin, C., Berthet, L., Le Moine, N., Lerat, J., Loumagne, C., Oudin, L., Mathevet, T., Ramos, M., and Valéry, A. (2009). Crash tests for a standardized evaluation of hydrological models. *Hydrology and Earth System Sciences Discussions*, (13):p–1757.

Anh, N. L., Boxall, J., Saul, A., and Willems, P. (2010). An evaluation of three lumped conceptual rainfall-runoff models at catchment scale. *Proceedings of the 3rd International Symposium on British Hydrological Society, Newcastle, UK, July.*

Bayliss, A., Black, K., Fava-Verde, A., and Kjeldsen, T. (2006). URBEXT2000 - a new FEH catchment descriptor. calculation, dissemination and application. Technical Report R&D FD1919/TR, Department for Environment Food and Rural Affairs, CEH wallingford.

Beven, K. J. (2011). *Rainfall-runoff modelling: the primer.* John Wiley & Sons, Chichester.

Bouffard, J.-S. (2014). *A comparison of conceptual rainfall-runoff modelling structures and approaches for hydrologic prediction in ungauged peatland basins of the James Bay lowlands.* PhD thesis, Carleton University Ottawa. Retrieved from https://curve.carleton.ca/7ed3b15f-bfff-4027-9a87-f314602d7a1a.

Clark, M. P., McMillan, H. K., Collins, D. B., Kavetski, D., and Woods, R. A. (2011). Hydrological field data from a modeller's perspective: Part 2: process-based evaluation of model hypotheses. *Hydrological Processes*, 25(4):523–543.

Donnelly-Makowecki, L. and Moore, R. (1999). Hierarchical testing of three rainfall–runoff models in small forested catchments. *Journal of Hydrology*, 219(3):136–152.

Duan, Q., Gupta, V. K., and Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Applications*, 76(3):501–521.

Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans*, volume 38. Society for Industrial and Applied Mathematics.

Euser, T., Winsemius, H., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., and Savenije, H. (2013). A framework to assess the realism of model structures using hydrological signatures. *Hydrology and Earth System Sciences*, 17(5):1893–1912.

Ewen, J. (2011). Hydrograph matching method for measuring model performance. *Journal of Hydrology*, 408(1):178–187.

Ewen, J. and O'Donnell, G. (2012). Prediction intervals for rainfall–runoff models: raw error method and split-sample validation. *Hydrology Research*, 43(5):637–648.

Gharari, S., Hrachowitz, M., Fenicia, F., and Savenije, H. (2013). An approach to identify time consistent model parameters: sub-period calibration. *Hydrology and Earth System Sciences, 17, 2013.*

Gupta, H. V., Wagener, T., and Liu, Y. (2008). Reconciling theory with observations: elements of a diagnostic approach to model evaluation. *Hydrological Processes: An International Journal*, 22(18):3802–3813.

Jones, D. A. and Kay, A. L. (2007). Uncertainty analysis for estimating flood frequencies for ungauged catchments using rainfall-runoff models. *Advances in Water Resources*, 30(5):1190–1204.

Keller, V., Tanguy, M., Prosdocimi, I., Terry, J., Hitt, O., Cole, S., Fry, M., Morris, D., and Dixon, H. (2015). CEH-GEAR: 1 km resolution daily and monthly areal rainfall estimates for the UK for hydrological and other applications. *Earth System Science Data*, 7(1):143–155.

Kjeldsen, T., Stewart, E., Packman, J., Folwell, S., and Bayliss, A. (2005). Revitalisation of the FSR/FEH rainfall-runoff method. Technical report, Defra R&D Technical Report FD1913/TR, CEH Wallingford.

Klemeš, V. (1986). Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, 31(1):13–24.

Krause, P., Boyle, D., and Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, 5:89–97.

Legates, D. R. and McCabe, G. J. (1999). Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, 35(1):233–241.

McMillan, H., Westerberg, I., and Branger, F. (2016). Five guidelines for selecting hydrological signatures. *Hydrological Processes*, 31(26):4757–4761.

McMillan, H. K., Clark, M. P., Bowden, W. B., Duncan, M., and Woods, R. A. (2011). Hydrological field data from a modeller's perspective: Part 1. diagnostic tests for model structure. *Hydrological Processes*, 25(4):511–522.

Mishra, S. (2009). Uncertainty and sensitivity analysis techniques for hydrologic modeling. *Journal of Hydroinformatics*, 11(3-4):282–296.

Nash, J. E. and Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I-A discussion of principles. *Journal of Hydrology*, 10(3):282–290.

Pappenberger, F. and Beven, K. J. (2006). Ignorance is bliss: Or seven reasons not to use uncertainty analysis. *Water Resources Research*, 42(W05302).

Pechlivanidis, I., Jackson, B., and McMillan, H. (2010). The use of entropy as a model diagnostic in rainfall-runoff modelling. In *iEMSs 2010. International Congress on Environmental Modelling and Software. 5 July, Ottawa, Canada*, volume 2, pages 1780–1787.

Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika*, 43(3/4):353–360.

Refsgaard, J. C. (1997). Parameterisation, calibration and validation of distributed hydrological models. *Journal of Hydrology*, 198(1):69–97.

Refsgaard, J. C. and Knudsen, J. (1996). Operational validation and intercomparison of different types of hydrological models. *Water Resources Research*, 32(7):2189–2202.

Robinson, E.L.Blyth, E. D.-P. E. J. A. (2016). Climate hydrology and ecology research support system potential evapotranspiration dataset for great britain (1961-2015). data retrieved from NERC Environmental Information Data Centre, `https://doi.org/10.5285/8baf805d-39ce-4dac-b224-c926ada353b7`.

Schaefli, B. and Gupta, H. V. (2007). Do Nash values have value? *Hydrological Processes*, 21(15):2075–2080.

Selle, B. and Hannah, M. (2010). A bootstrap approach to assess parameter uncertainty in simple catchment models. *Environmental Modelling & Software*, 25(8):919–926.

Van Belle, G. (2011). *Statistical rules of thumb*, volume 699. John Wiley & Sons.

Vogel, R. M. and Sankarasubramanian, A. (2003). Validation of a watershed model without calibration. *Water Resources Research*, 39(10).

Wagener, T., Sivapalan, M., Troch, P., and Woods, R. (2007). Catchment classification and hydrologic similarity. *Geography compass*, 1(4):901–931.

Weglarczyk, S. (1998). The interdependence and applicability of some statistical quality measures for hydrological models. *Journal of Hydrology*, 206(1-2):98–103.

Willmott, C. J. and Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1):79–82.

The previous chapter (4) compared both URMOD and DAYMOD on 28 urban catchments in the Thames catchment. The purpose of the chapter was to answer Research questions 1 and 3. The results show that simply implementing a simple urban runoff framework and routing methodology does improve model performance but further study is needed in order to answer research question 1. The answer to **Research question 3:** Can simple statistical-based techniques be used to improve validation of rainfall-runoff models. was answered during this chapter, of the two hydrological model comparison tools developed in Chapter 4, the binomial hypothesis test is used again in the next two chapters. Whilst the paired Z-test is not, as outlined in the conclusion in Chapter 4 further research is needed in order to determine if it is an effective tool to compare model performance. However the jackknife calibration/validation method will again be used in the next two chapters. The next chapter compares URMOD and DAYMOD but URMOD will use the parallel linear reservoir detailed in Chapter 3.2.5. In order to answer research question 1 a detailed analysis of two catchments using sub-daily data. The purpose for this as outlined in the introduction of Chapter 5 is because of the nature of urban surfaces urban flow within a catchment may have left within a day, and so to test the performance of the URMOD model a comparative analysis on sub-daily data will be conducted. The table below outlines the contribution of the thesis author and the co-author of the coming paper chapter.

<table>
<tr><td colspan="10">This declaration concerns the article entitled:</td></tr>
<tr><td colspan="10">Exploring the impact of urbanisation on lumped rainfall-runoff modelling using sub-daily data</td></tr>
<tr><td colspan="10">Publication status (tick one)</td></tr>
<tr><td>Draft</td><td>✓</td><td>Submitted</td><td></td><td>Review</td><td></td><td>Accepted</td><td></td><td>Published</td><td></td></tr>
<tr><td colspan="2">Publication details (reference)</td><td colspan="8">James Fidal, Thomas Kjeldsen. Draft format, target journal: Journal of Hydrology. Submission date expected April 2019.</td></tr>
<tr><td colspan="2">Candidate's contribution to the paper (detailed, and also given as a percentage).</td><td colspan="8">The author of this thesis contributed to the entire paper, formulating the ideas, defining the methodology used, application of the models and preparation/ writing of the manuscript. Data collection was not handled as part of this paper, relevant data collectors acknowledged in the paper. Each author's exact contribution is detailed below

Formulation of ideas: 90% James, 10% Thomas
Model design and coding: 80% James, 20% Thomas
Model application: 100% James
Analyse of model output: 100% James
Preparation of manuscript 80% James 20% Thomas</td></tr>
<tr><td colspan="2">Statement from Candidate</td><td colspan="8">This paper reports on original research I conducted during the period of my Higher Degree by Research Candidature.</td></tr>
<tr><td colspan="2">Signed</td><td colspan="2">J.Fidal</td><td colspan="2">Date</td><td colspan="4">October 3 2018</td></tr>
</table>

Table 4.3: Statement of authorship paper:
Operational model comparison techniques for rainfall-runoff models.

Chapter 5

Exploring the impact of urbanisation on lumped rainfall-runoff modelling using sub-daily data

5.1 Abstract

The introduction of urban land-cover changes the hydrological characteristics of catchments. Hence there is a need to develop new models to account for this change in urbanisation. The issue that is addressed in this study is assessing the performance of an urban rainfall-runoff model, when sub-daily data is applied. This is achieved through comparing two rainfall-runoff models, a full urban rainfall runoff with urban pipe routing and a rural model. These models are applied to two urban catchments in the Thames river catchment, the river Cut and river Ray catchments. The results show that for the river Cut the full urban model outperformed the rural model, whilst for the river Ray neither model dominated the other.

5.2 Introduction

It is well documented that urbanisation has a detectable impact on the hydrology of a catchment. This impact ranges from increased runoff rates in catchments (Fletcher et al. (2013) and Jones et al. (2000)) to an increase in peak flows in catchments (Miller et al. (2014) and Rose and Peters (2001)). Alongside the increase in runoff generated from catchments the time in which runoff leads to the catchment outlet is shortened, with Shaw (1994) concluding that the lag time between precipitation and river runoff is decreased. One method that is used to capture the effects of urbanisation on catchment hydrology is rainfall-runoff modelling.

Rainfall-runoff models that explicitly account for urbanisation play an important role in simulating the effects of urban land-use by representing the effects of impervious surfaces and other urban systems. From the published literature it is evident that general urban processes, besides simple representations of impervious areas, are rarely considered and implemented into rainfall-runoff models. For example, Redfern et al. (2016), Fletcher et al. (2013) and Jacobson (2011), discussed that the complex nature of urban environments, most rainfall-runoff models are rarely parametrised to fully account for urban processes other than just implementing an impervious surface extension; with 30% of the models in the review presented by Salvadore et al. (2015) only accounting for the impact of impervious cover on infiltration. This issue is increasingly complex when determining which of the urban processes have the largest impact on the hydrological response, and as such raises the issue of generalising these urban processes for rainfall-runoff modelling. One such urban process is urban flow routing. Urban routing on a daily time step is potentially ineffective, with Mitchell et al. (2001) concluding that excess rainfall will have left most urban catchments in a matter of hours. Hence urban flow routing is unlikely to be effective on a daily time scale. Figure 3 in Salvadore et al. (2015) shows the spatial and temporal scales of hydrological processes from a number of studies, indicating that sewers and storm drainage lag-times are over a day long for catchments over 100km.

As the hydrological cycle within catchments becomes more complex due to urbanisation, new models have been developed to try to account for these effects, thereby becoming more complex which can lead to over-parameterisation.

Loague and Freeze (1985) highlighted the problem of over-parameterisation and concluded that simpler models applied to a number of catchments generally performed better than more complex physically based models. This finding was supported by Jakeman and Hornberger (1993) who concluded that the accuracy of model parameters can be limited by the amount or level of accuracy of data which is used.

Similarly Orth et al. (2015) compared three models with increasing complexity across multiple catchments in Switzerland, concluding that the performance of the models is influenced by the hydrological conditions, such that simpler models can out-perform more complex models. Perrin et al. (2001) also discussed the problem of model complexity, through a study compared 19 model structures with three to nine optimised parameters on 429 catchments, mostly in France but also including catchments in United States, Australia, Ivory Coast and Brazil. They conclude that the simpler models (three parameters) can perform as well as the nine parameter models, with more complex models often showing signs of over-parameterisation.

While all of these studies reviewed above compare individual models, another possible approach is to improve upon an existing model structure. Butts et al. (2004), used five different model performance criteria to compare ten different variations in model structure of the MIKE SHE and MIKE 11 models. They found large variations in model performance and suggests that using a combination of model results rather than a single version of any one model. Similarly Li et al. (2015) used four different model performance criteria to assess the performance of five versions of the HBV model with increasing complexity when applied to the 18,932 km^2 Norsfoss catchment in Norway. Results showed that increasing model complexity improved the runoff simulation but the performance of more complex models could not be distinguished from each other. This indicates that whilst increasing the model complexity can improve model performance this may not necessary lead to improved accuracy of the model.

These studies show how a model structure can be improved to create a new version of a model an alternative approach. The approach taken by Butts et al. (2004) was to expand upon two existing models in order to create a multi-model setup, the new models created were nested model structures of the original model. A nested model is a model structure in which the processes and parameters are a subset of a more complex model with more parameters and processes. This paper will take a similar approach and present a new urban extension framework for implementing the urban water cycle into a lumped conceptual rural rainfall-runoff model, accounting for both urban runoff generation process and urban routing, thereby creating a nested model structure. This model is named the URMOD model, and features a nested model structure so when no urbanisation is present the model reverts back to the rural model, which is based on the already existing DAYMOD model as described by Packman (2004) and Kjeldsen et al. (2005). This study introduces a new approach where, in line with findings from experimental studies, infiltration through urban surfaces depends on the soil moisture of the underlying soil and a separate urban scaling factor.

5.3 Model development

This study develops a deterministic, continuous-time, lumped, conceptual rainfall-runoff model for simulating in catchments with a large proportion of urban land-use (URMOD). The model is structured so that a catchment is split into a rural and an urban section, with each section having different infiltration (and thus runoff) and routing characteristics. The URMOD model was designed to be applicable to both rural and urban catchments, so that for catchments with no urbanisation URMOD defaults to a completely rural model. The rural section of the model is an adaptation of an existing model described in Packman (2004) forming part of the ReFH model routinely used for design flood estimation in the UK (Kjeldsen et al., 2005).

The new urban section of URMOD provides a conceptual representation of sealed urban areas and routing which employs a small amount of parameters and is therefore parsimonious. URMOD has nine parameters in need of calibration (DAYMOD has seven calibrated parameters). Four of DAYMODs parameters are used for runoff generation and the other three for flow routing. These seven parameters are included in URMOD and the extra two parameters are used for the urban processes, one for urban runoff generation and the second for urban routing. URMODs structure consists of two main processes; (i) the soil column representing infiltration and runoff generation (Section 5.3.1 and 5.3.2), and (ii) urban, base flow and surface model routing (Section 5.3.3 and 5.3.3). For consistency in the rest of the paper the two models will be labeled as follows, the full urban model URMOD_1 and the default rural model URMOD_2.

5.3.1 Infiltration and runoff model without urbanisation

Surface runoff (κ) and infiltration (η) are modelled using a soil-column based approach. Where precipitation (i) that does not infiltrate into the soil column is converted into direct runoff. Hence a simple mass balance can be defined rain=runoff+infiltration ($i = \kappa + \eta$). The fraction of precipitation that is turned to either runoff or infiltrates depends on the soil moisture level but differs depending on the rural or urban section of the model. The rest of this section will describe the infiltration model for use in rural areas, with section 5.3.2 presenting the urban infiltration model.

The infiltration and runoff generation in URMOD is based on a Probability Distributed Model (PDM) Moore (1985), with uniformly distributed soil moisture capacity. The PDM assumes that the soil moisture capacity (C) varies randomly over the entire catchment between a value of zero and C_{max}, according to a uniform distribution, such that capacities occur with equal frequency. An initial moisture content C_0 is chosen (full saturation is assumed for $C < C_0$), runoff is generated from areas with a capacity less than C_0, whereas the other areas are unsaturated $(C - C_0) > 0$ and no runoff is generated. Since soil moisture capacity is uniformly distributed and the maximum of the soil moisture capacity is denoted C_{max}, the mean of C equals the mean soil moisture capacity (S) in the catchment and is defined as,

$$S = \frac{1}{2} C_{max}. \tag{5.1}$$

For details on the derivation of the infiltration equation Eq 5.2 see Section 3.1.1. The infiltration equation is defined in Eq 5.2

$$f = \left(1 - \frac{m}{S}\right)^{\frac{1}{2}}. \tag{5.2}$$

Let m represent the soil moisture at time t, such that $\frac{m}{S}$ denotes the soil columns saturation level. If $\frac{m}{S} = 1$ the soil column is completely saturated and $\kappa = i$ (100% runoff, 0% infiltration) whereas if the soil is completely dry $\frac{m}{S} = 0$ all rain infiltrates, $\eta = i$ (0% runoff, 100% infiltration). Whilst this is a oversimplication when the soil column is dry, it is expected the column will not be completely dry except for extreme periods of no rain. Since it is assumed that the precipitation that does not infiltrate into the ground turns into runoff, the runoff equation can be defined as:

$$q = i - f = i\left[1 - \left(1 - \frac{m}{S}\right)^{\frac{1}{2}}\right]. \tag{5.3}$$

Eq 5.2 and Eq 5.3 are used to describe the amount of rainfall that is turned into direct runoff or infiltrates into the soil. For η and κ to be estimated, the moisture content m needs to be accounted for in the water balance of the soil column. The change of water in the soil is driven by three processes: (i) infiltration into the soil, (ii) drainage out of the soil and (iii) evaporation of water back to the atmosphere. The soil column is assumed to have three different zones depending on soil moisture levels which will change the drainage and the evaporation as shown by the soil column on the in Figure 1.

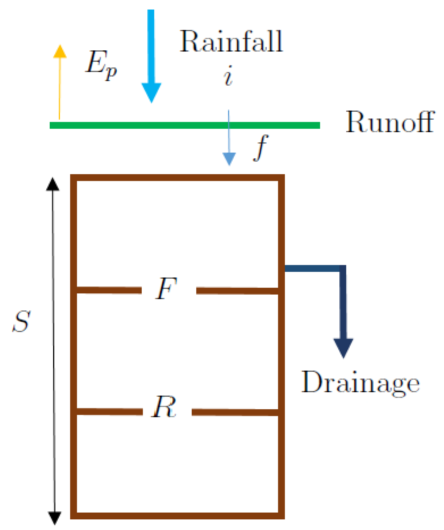


Figure 5-1: Conceptual soil column from DAYMOD

Zone 1, near the soil surface, is defined as when the soil moisture is above the field capacity ($m > F$). In this case the evaporation is assumed at the potential rate (E_p) and drainage depends upon moisture content (m) and a calibrated drainage coefficient (k) so drainage out of the column takes place at a rate of $k(m - F)$. Zone 2 is when the soil moisture does not exceed the field capacity but does exceed the rooting depth ($R < m < F$) the evaporation is again at the potential rate (E_p) and there is no drainage. Zone 3 when soil moisture is below the rooting depth ($m < R$), there is again no drainage and evaporation reduces linearly with depth as $E = E_p \frac{m}{R}$ until it reaches $E = 0$ for $m = 0$. Three different equations have been developed to govern the soil moisture dynamics of each zone. The infiltration term in each of these equations does not change and is determined by Eq 5.2. The change in soil moisture at a given time t when moisture exceeds the field capacity ($m > F$) is presented in Eq 5.4. Eq 5.4 is solved using a finite difference method which is detailed in Section 3.1.1 :

$$\frac{dm}{dt} = \underbrace{i\left(1 - \frac{m}{S}\right)^{\frac{1}{2}}}_{\text{Infiltration}} - \underbrace{k(m - F)}_{\text{Drainage}} - \underbrace{E_p}_{\text{Evaporation}} . \tag{5.4}$$

In Eq 5.4 the drainage and evaporation terms will change depending on the zone in which the level of moisture is located. If the soil moisture changes zone mid time step, the time step is divided and the remaining period uses the new zone.

5.3.2 Infiltration across urban areas

If urban land-use is present then infiltration across the catchment will be made up of two contributions representing the rural areas and the urban areas, respectively. As URMOD is a lumped model, the total infiltration is represented as a weighted average of infiltration on the two land-use types:

$$f = i(1 - u)f_{rur} + iuf_{urb}, \tag{5.5}$$

where u is the percentage of the total catchment area covered by urban land-use, f_{rur} represents infiltration in the rural areas as defined in Eq 5.2 and infiltration in urban areas is denoted f_{urb}.

The infiltration in urban areas is dependent on soil moisture, but is reduced by a parameter γ. The γ parameter accounts for the variability of infiltration in different urban areas, such that as γ increases less infiltration occurs in urban areas. f_{urb} is defined as:

$$f_{urb} = i\left[(1-\gamma)\left(1-\frac{m}{S}\right)^{\frac{1}{2}}\right]. \tag{5.6}$$

If $\gamma = 0$ then infiltration for the urban area is the same as the infiltration for the rural area, and if $\gamma = 1$ then the urban area would be completely sealed and there would be no infiltration. γ is a calibrated parameter introduced in an attempt to account for variability of infiltration across different urban surfaces. Substituting the rural infiltration Eq 5.2 and the urban infiltration Eq 5.6 into Eq 5.5 gives the total infiltration as:

$$f = \underbrace{i(1-u)\left(1-\frac{m}{S}\right)^{\frac{1}{2}}}_{\text{rural}} + \underbrace{iu(1-\gamma)\left(1-\frac{m}{S}\right)^{\frac{1}{2}}}_{\text{urban}}.. \tag{5.7}$$

Next, replacing the infiltration term in the soil moisture accounting model (Eq 5.4) with Eq 5.7 gives the soil column water balance for urban catchments as:

$$\frac{dm}{dt} = \underbrace{i(1-u)\left(1-\frac{m}{S}\right)^{\frac{1}{2}} + iu(1-\gamma)\left(1-\frac{m}{S}\right)^{\frac{1}{2}}}_{\text{Infiltration}} - \underbrace{k(m-F)}_{\text{Drainage}} - \underbrace{E_p}_{\text{Evaporation}}. \tag{5.8}$$

Similar to Section 5.3.1 there are three zones in the soil column, hence Eq 5.8 is the model of soil moisture with respect to a time step t when soil moisture is exceeding the field capacity. Similar to Eq 5.4 the drainage and evaporation will change depending on the zone. The model in Equation 5.7 predicts that infiltration is reduced as a function of both fraction of urbanisation and increasing soil moisture. The percentage of precipitation that does infiltrates is defined as a function of soil moisture (Equation 5.7) as shown in Figure 5-2 for a fixed value of $\gamma = 0.7$

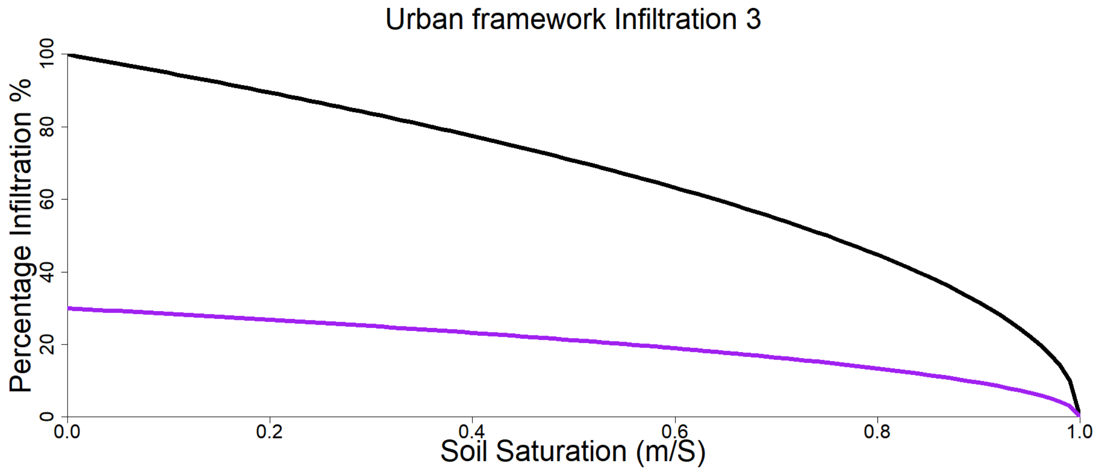


Figure 5-2: Percentage infiltration versus soil saturation. Urban framework 3: Urban infiltration (purple) $\gamma = 0.7$, Rural infiltration (Black)

5.3.3 Routing model

URMOD applies separate routing of the runoff generated from the rural and urban parts of catchments as shown in Figure 5-2. The runoff generated from the rural parts of the catchment is split and a contribution goes to the baseflow and a contribution goes to the surface flow. The proportion of the runoff that contributes to the baseflow is first routed through a local baseflow reservoir with lag B_L, before it emerges into the channel and is then routed through a channel linear reservoir of lag S_L in order to obtain the baseflow at catchment outlet. Whilst the proportion of runoff which contributes to the surface flow is only routed through the channel linear reservoir, before combining with the baseflow for the rural flow at the catchment outlet. The runoff generated from urban parts of the catchment is routed through a separate linear reservoir of lag U_L before it reaches the catchment outlet, where it is combined with the rural flow to generate the total flow (q_{sim}) at the catchment outlet.

Single linear reservoir routing model

The routing method for the rural section of URMOD is based on the linear reservoir concept, with a characteristic recession defined as an exponential decay. Two different lag times are used, B_L for the base flow lag and S_L for surface flow lag. The routing in URMOD assumes that the areas of the catchment that produces the surface runoff also produces the baseflow. The base flow recharge, r, that feeds a linear reservoir, with storage σ_B is given as B_L times the outflow b. As shown in Eq 5.9:

$$\sigma_B = B_L b. \tag{5.9}$$

Hence the change in storage can be determined as:

$$\frac{d\sigma_B}{dt} = r - b. \tag{5.10}$$

The continuity equation (Eq 5.10) can be combined with Eq 5.9 and solved over time step δt spanning $[0, t]$ this gives the local base flow rate Eq 5.11:

$$b_t = b_0 e^{\frac{-t}{B_L}} + r\left(1 - e^{\frac{-t}{B_L}}\right). \tag{5.11}$$

The average baseflow over the time-step can then be obtained as $((b_0 + b_t)/2)$. Next the local baseflow b_t is routed though the channel routing model, which is obtained using the same method used to obtain Eq 5.11 with surface flow lag S_L and surface flow outflow s:

$$s_t = s_0 e^{\frac{-t}{S_L}} + z(1 - e^{\frac{-t}{S_L}}). \tag{5.12}$$

The proportion of the total runoff designated as surface runoff is only routed through Eq 5.12, such that $(s_0 + s_t/2)$ is the average surface flow at the catchment outlet. Combining the baseflow (b_t) and surface flow (s_t) gives the total flow at the catchment outlet $(q_{sim} = b_t + s_t)$.

Parallel urban routing model

In this method the contribution of runoff from the urban areas is routed directly to the outlet via a separate and parallel linear reservoir. It is assumed that the urban area is one lumped entity transporting runoff to the outlet quicker than the runoff from the surrounding rural areas. This is done by defining an upper bounded linear reservoir representing the convergence in storm water pipes and which has a lag of U_L, the upper limit of the bounded pipe is set at 1 m^3/s. Whilst the linear reservoirs in the rural area does not have an upper capacity, the upper bounded nature of the pipe system meaning that if the pipe system reaches the full capacity the extra runoff spills over to the rural part of the catchment and thus contributes to the next time step's rural routing (before the runoff is designated as baseflow or surface flow). This is an attempt to simulate the finite capacity of the pipes in the urban drainage network. The solution to the linear storage equation will be used to determine the urban routing. Let σ_u be the storage of the pipe system with U_L being the lag time and v being the outflow of the system:

$$\sigma_p = U_L v. \tag{5.13}$$

The change in the storage equation for Equation 5.13 can be solved via numerical methods as opposed to analytical integration which gives Equation 5.14. The derivation is detailed in Section 3.2.5 :

$$v_t = \frac{2U_L v_0 + \Delta t(z_t - v_0 + z_0)}{2(U_L + \Delta t)}, \tag{5.14}$$

v_t is then combined with the total surface flow and base flow from the rural section of the model to generate the flow at the catchment outlet denoted q_{sim}. A full visual representation of URMOD is displayed in Figure 5-3

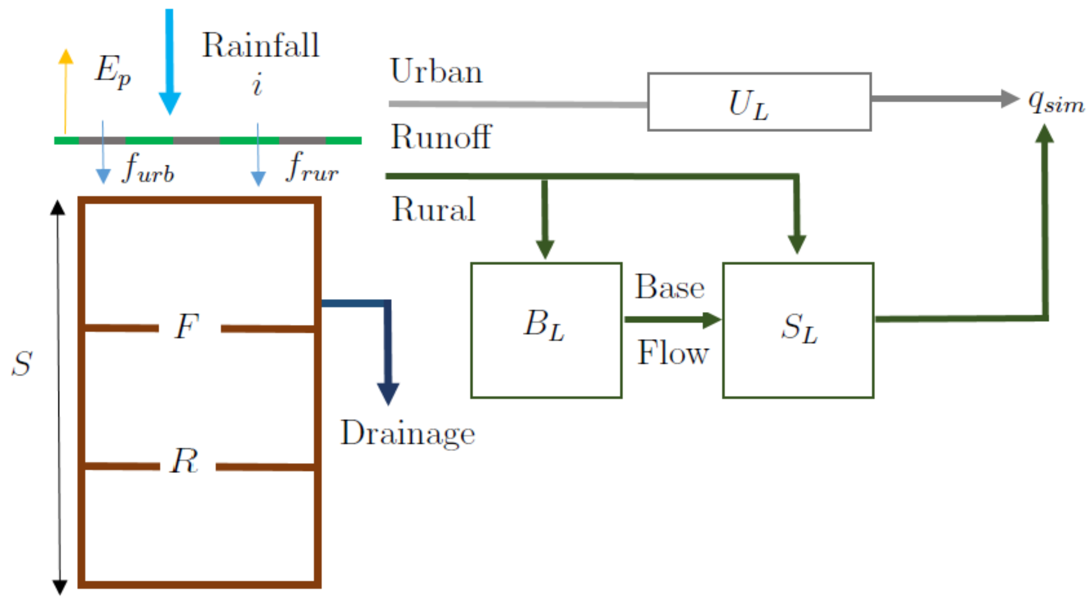


Figure 5-3: Visual representation of the URMOD model

Since all of the default rural model (DAYMOD) processes are included within the full urban model structure, DAYMOD is a nested model of URMOD. Two variations of the nested model structure will be employed in this experiment (i) the full urban model $URMOD_1$ and (ii) The default rural model without urban components $URMOD_2$ (DAYMOD). For reference all of the inputs, outputs and parameters for $URMOD_1$ and $URMOD_2$ are found in Table 1 in Appendix 5.A.

5.4 Case Study- Cut and Ray catchments

Two catchments, each with a high fraction of urban land-cover were chosen for this case study. The first is the River Ray catchment (NRFA number 39087) containing the town of Swindon. The second the River Cut catchment (NRFA number 39052) containing the town of Bracknell. Both catchments are located within the Thames, basin catchment as shown on Figure 5-4.

Bracknell is a town located in the county of Berkshire approximately 30 miles west of London. It has a population of approximately 50,000. The area of the catchment is 50.2 km^2 with the town of Bracknell covering approximately 24.13% of the catchment in 2010 which has increased from 4.61% from the 60s (Putro et al., 2016). The undeveloped areas is primarily forest underlain with London clay.

Swindon is a town located in Wiltshire 78 miles west of London, with a population of approximately 182,441. The area of the catchment is 84.1 km^2 with the town of Swindon covering approximately 22% of the catchment in 2010 Putro et al. (2016). The gauging station uses a ultrasonic flow gauge in order to measure flow. The undeveloped area is primarily agricultural.

5.4.1 Hydro-meteorological data

The hydro-meteorological data for both catchments consists of: observed average precipitation (i), observed average river flow (q_{obs}) and potential evaporation data (E_p). Data with two different levels of temporal resolution were used: (i) daily average data and (ii) sub-daily data (hourly). The daily precipitation data were obtained from the CEH GEAR data set spanning a 50 year period (1961-2012) Keller et al. (2015). The evaporation data was obtained from the Climate,

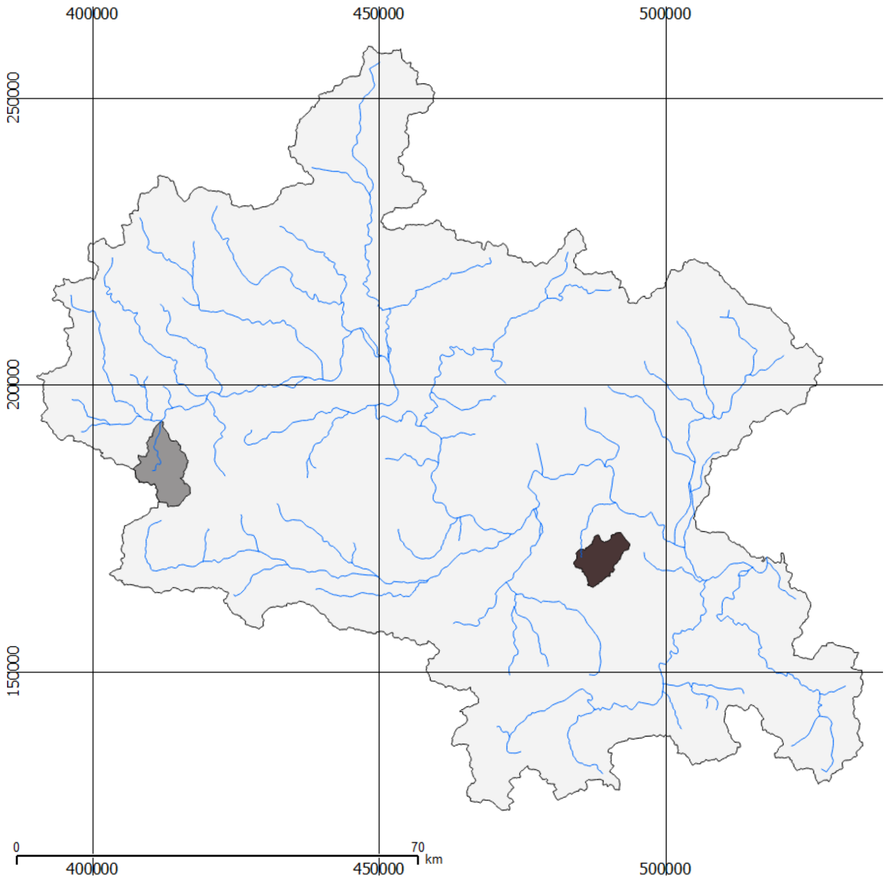


Figure 5-4: Map of Thames river catchment with Ray (left in light grey) and Cut (right in dark grey) catchments highlighted.

Hydrology and Ecology research Support System (CHESS) (Robinson, 2016). The daily river flow data for both catchments were acquired from the National River Flow Archive (NRFA) spanning a 38 year period (1975-2013). The daily rainfall and evaporation daily data had no missing observations, where as the flow data from the Ray catchment (NRFA: 39087) had some missing observations.

The sub-daily data was obtained in hourly time steps, and converted into four, eight and twelve hourly time step for the purpose of this study. The hourly sub-daily data was obtained as apart of the POLL-CURB project (Miller, James, (pers.comm.) 2017). The sub-daily flow data spanned 20 years (1990-2010) and similar to the daily data there was missing observations. When the data was converted to the relevant time steps if the missing observation was part of the average then the particular data point was removed. The sub-daily precipitation data spanned 20 years (1990-2010) and again had missing observations. Similar to the flow data the missing observations were removed once converted. No sub-daily evaporation values were obtained so averaged values from the disaggregated daily data were used.

One important input for the urban model is determining the percentage of urbanisation in a catchment. The URBEXT_{2000} catchment descriptor Bayliss et al. (2006) was used for this study, where the subscript 2000 denoting that the 50m $\times$ 50m land-cover data that was used to construct the index refers to the period between the years of 1998-2000. URBEXT_{2000} uses a contribution of both urban and sub-urban land-cover classes. Only half of the sub-urban area are counted as urban as it is assumed that half of the sub-urban area is made up of vegetation e.g (gardens or parks) Bayliss et al. (2006). URBEXT values for 1990 were obtained for river Ray and Cut catchments, these are calculated similar to the year 2000 value, but using land use data from 1990.

5.4.2 Model Calibration and Validation

Calibration of URMOD's model parameters requires access to good quality data consisting of long-term series of rainfall, runoff and evaporation data at a common time step. Calibration of the nine free model parameters (seven for DAYMOD) is achieved via minimising an objective function. The algorithm used to minimise the objective function is the shuffled complex evolution (SCE) approach, further details of the SCE are provided in Duan et al. (1993). In order to calibrate the model parameters suitable calibration and validation periods need to be defined. Traditional methods of model calibration use the split-sample method defined by Klemeš (1986), in which the entire data set is split into two non over-lapping calibration and validation period. The method used in this study is an extension to split-sample method which, rather than splitting the entire data set into two periods, splits the data set into a number of subsets of equal length and will systematically calibrate the model on each subset. This methodology is called the Jackknife approach, and is an adaptation of the methodology used by Jones and Kay (2007) in which the jackknife method is used to quantify model parameter uncertainty. The jackknife method is used in this paper to obtain the uncertainty of model performance criteria for the purpose of model comparison.

5.4.3 Jackknife calibration methodology

Sub-daily and daily data covering a 20-year period were used for this study, ranging from 1990-2009. This period was chosen to make use of the two URBEXT criteria obtained from the Flood estimation handbook (of Hydrology, 1999), URBEXT_{2000} and URBEXT_{1990}. A calibration length of 5 years was chosen for the daily data. The remaining 15 years will be split into three 5 year continuous periods, to obtain three validation sub periods. A calibration length of 2 years was chosen for the sub-daily data. The remaining 18 years will be split into nine 2 year continuous sub periods, to obtain nine validation sub periods per year. This method is presented in Figure 5-5.

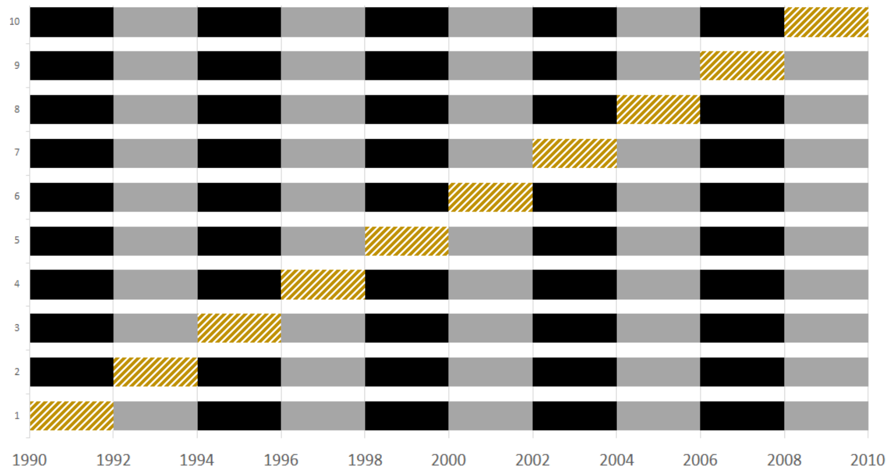


Figure 5-5: 20 year jackknife method, calibration period in hatched gold. The grey and black boxes show the various 2-year calibration periods.

The Jackknife calibration method is presented for the sub-daily data, considering the 20 years of data, starting at the first year 1990. A sub period consisting of 2 years of continuous data are omitted, and the model parameters are calibrated on the removed 2 years of data obtaining a parameter set θ_1. The parameter set is applied to URMOD in order to obtain simulated runoff for the remaining 18 years. Once completed the omitted 2 years of data is put back into the data set and the next 2 years of continuous data are omitted and calibrated to generate another parameter set θ_2. The parameter set is applied to URMOD in order to obtain simulated runoff for the remaining 18 years of data. Using this procedure performance criteria can be obtained for every 2 years of data by comparing simulated runoff and observed runoff to obtain 9 performance criteria. This process is repeated systematically to obtain 10 parameter sets $\theta_j, j = 1, ..., 10$ and 9 performance criteria for each 2 year period $Z_{j,k}, j = 1, ..., 10, k = 1, ..., 9$. This jackknife process is applied to all four data aggregations (hourly, 4-hourly, 8-hourly and 12-hourly) for both catchments. A similar process is used for the daily data but a calibration period of 5 years is used. This methodology will be applied to both URMOD_1 and URMOD_2 to obtain $Z_{1,j,k}$ and $Z_{2,j,k}$, the subscript 1 and 2 denote which model performance criteria is obtained. The difference in performance of both models ($c = j \times k = 90$) performance criteria can be obtained.

$$Z_{d,j,k} = Z_{1,j,k} - Z_{2,j,k}, j = 1, ..., 10, k = 1, ..., 9. \tag{5.15}$$

The mean value for each 2-year period can be obtained.

$$\overline{Z_{d,k}} = 10^{-1} \sum_{j=1}^{10} Z_{d,j,k}. \tag{5.16}$$

The difference in performance criteria $Z_{d,j,k}$ will be applied to a binomial approach detailed in Section 5.4.5, in order to explore the models performance.

5.4.4 Model performance criteria

Two performance criteria are adopted for this study: the Nash-Sutcliffe efficiency statistic NSE (Nash and Sutcliffe, 1970) and the mean absolute error MAE. The NSE is defined as :

$$NSE = 1 - \frac{\sum_{t=1}^{n} (q_{obs} - q_{sim})^2}{\sum_{t=1}^{n} (q_{obs} - \bar{q}_{obs})^2}. \tag{5.17}$$

Where q_{sim} is the simulated river flow from URMOD, q_{obs} is the observed river flow and $\bar{q}_{obs}$ denoting the mean of the observed river flow. Values of NSE lie between one and $-\infty$, with a value of one indicating perfect fit, i.e $q_{obs} = q_{sim}$. The second criteria is the mean absolute error MAE defined as

$$MAE = \frac{\sum_{t=1}^{n} |q_{obs} - q_{sim}|}{n}. \tag{5.18}$$

Values of MAE are bounded by 0 and ∞. Whereas for NSE, one is the perfect fit, a perfect fit will result in a MAE value of zero.

5.4.5 Binomial hypothesis test

The Binomial hypothesis test approach uses a success/failure approach to compare two models; either a model outperforms the other or it does not. The case where model performance is exactly equal is not taken into account, as this case is considered very unlikely occurrence. This method will be used to explore the performance of the models on the sub-daily data at each data length (hourly, 4-hourly, 8-hourly and 12-hourly) for both catchments. The 90 performance criteria differences $Z_{d,j,k}$ obtained from the jackknife method from both URMOD_1 and URMOD_2 is used. Let c denote the number of trails, such that a trail is defined as the difference in performance criteria $Z_{d,k}$ for a particular 2 year period the URMOD_1 and URMOD_2. The hypothesis test can be set up such that the probability that there is no difference in model performance, and therefore the probability that either model outperforms the other is assumed to be half, i.e:

$$H_0 : p = 0.5. \tag{5.19}$$

The alternative hypothesis (H_1) will be set up as a two tailed test, in order to test if either model's performance is significantly better than the other. The two tailed test is set up such that a success is defined as $Z_{d,j,k} > 0$, whereas a failure is defined as $Z_{d,j,k} < 0$. Hence the alternative hypothesis is given as:

$$H_1 : p \neq 0.5. \tag{5.20}$$

So let V be a random variable defined as the number of comparisons between $Z_{d,j,k} > 0$ or $Z_{d,j,k} < 0$. Thus the probability of v instances where $Z_{d,j,k} > 0$ is a binomial distribution $B(c, p)$ and given as;

$$P\{V = v\} = \binom{c}{v} \times p^v \times (1-p)^{(c-v)}. \tag{5.21}$$

A hypothesis test can be formed, for a predefined significance level, e.g $\alpha =$ 5%. The observed number of success v is compared to the critical interval defined as $v < B(c,p)_{\frac{\alpha}{2}}$ and $v > B(c,p)_{1-\frac{\alpha}{2}}$. With $\alpha/2$ and $(1-\alpha/2)$ denoting quantiles of the binomial distribution. Hence if v falls within the critical interval then the null hypothesis can be rejected such that there is a difference between the models. If v does not fall within the critical interval then the null hypotesis can be accepted such that there is no difference between the model performance.

5.5 Results

The results section is split into two main sections. The first section will focus on comparing the urban model (URMOD_1) and the rural model (URMOD_2) for the Ray catchment (NRFA 39087), exploring difference in performance criteria, the binomial hypothesis test and selected rainfall events presented as hydrographs. The second section will present the same analysis but for for the River Cut catchment (NRFA 39052).

5.5.1 Assessing performance of URMOD_1 and URMOD_2 on the River Ray catchment using performance criteria

This section explores the difference in performance of the URMOD_1 and URMOD_2 models on the River Ray catchment. Comparing performance on the daily time scale first, then exploring model performance on 12-hourly, 8-hourly, 4-hourly and hourly time aggregated data, respectively.

Daily comparison of URMOD_1 and URMOD_2

Figure 5-6 shows the difference in performance when using the NSE and MAE for the daily data. The left hand figure (a) is the difference in NSE, reported as $Z_{1,j,k} - Z_{2,j,k}$ and the right hand figure (b) is the difference in MAE, it is reported as $Z_{2,j,k} - Z_{1,j,k}$. Positive values indicate that URMOD_1 performs better than URMOD_2, indicated on Figure 5-6 as black circles. In contrast, negative values indicate that URMOD_2 performs better than URMOD_1, which are not represented on Figure 5-6 since no instances occurred. Performance criteria were obtained for the four calibration periods for both models, the difference of the calibration results are presented on Figure 5-6 as green squares. Lastly the numbers next to the points indicate which calibration period they are from, hence 1 indicates the calibration square's parameters were used for the rest of the points with a 1 next to them.

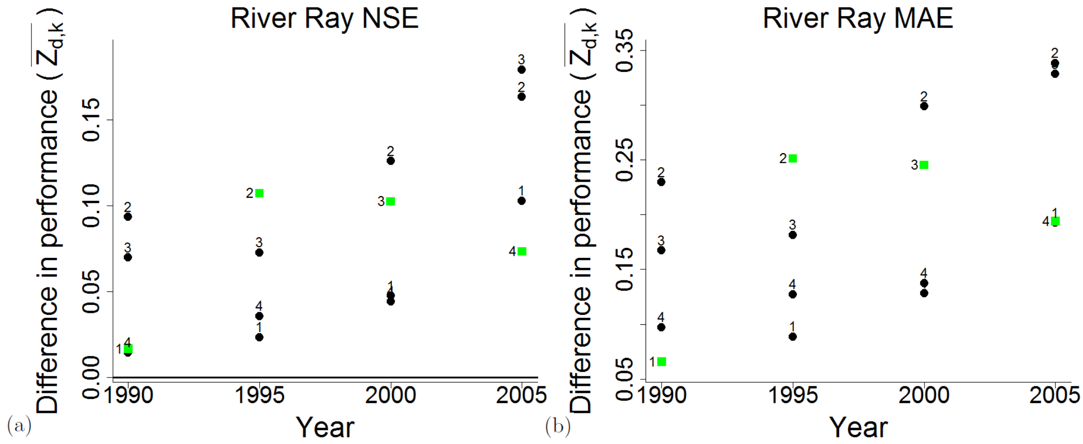


Figure 5-6: Comparison of models utilising daily data on river Ray using NSE criteria on the left (a) and MAE on the right (b).

The results in Figure 5-6 show that $URMOD_1$ (the complete urban model) has outperformed $URMOD_2$ (the rural model) for every validation period. Even though $URMOD_1$ has out performed $URMOD_2$ the values of NSE and MAE obtained for $URMOD_2$ are still considered very good. The average NSE performance of $URMOD_1$ for the 12 validation periods is 0.806, whilst average NSE for $URMOD_2$ was 0.725. The range of NSE values for $URMOD_1$ was 0.07, whereas $URMOD_2$ was 0.16. This indicates that $URMOD_1$ is a little more consistent than $URMOD_2$. The average MAE performance of $URMOD_1$ for the 12 validation periods is 0.404, whilst average MAE for $URMOD_2$ was 0.597. However the range of values was larger than that of the NSE, with $URMOD_1$ having a range of 0.148, and $URMOD_2$ 0.3. As only 12 validation periods (hence 12 data points) were used the Binomial hypothesis test was not applied. Whilst small sample sizes can be used for the Binomial hypothesis test this can be inaccurate due to the large variance from a small sample size.

Sub-daily comparison of $URMOD_1$ and $URMOD_2$

The difference in performance of the $URMOD_1$ and $URMOD_2$ models using 12-hourly, 8-hourly, 4-hourly and hourly data is presented Figures 5-7, 5-8, 5-9, 5-10. Due to 100 data points (90 validation and 10 calibration) being obtained, it is infeasible to present every comparison on a single graph. So an average of each 2 year validation period is obtained and the difference in performance criteria presented. The results shown in the left hand figures (a) for the NSE are presented such that a positive difference in average criteria ($\overline{Z_{d,k}} > 0$) indicates the URMOD$_1$ performance better than URMOD$_2$ shown as black circles on the figures. In contrast a negative difference ($\overline{Z_{d,k}} < 0$) indicates that the URMOD$_2$ model performed better than URMOD$_1$ shown as the red triangles. The right hand figures (b) show the MAE performance, with the same circle and triangle conclusion.

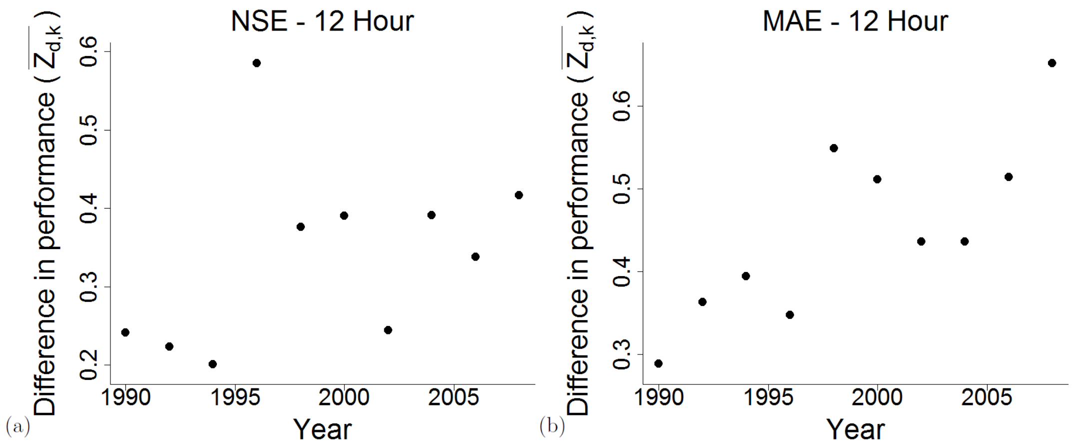


Figure 5-7: Comparison of models using 12-hourly data on River Ray catchment using NSE criteria on the left (a) and MAE on the right (b).

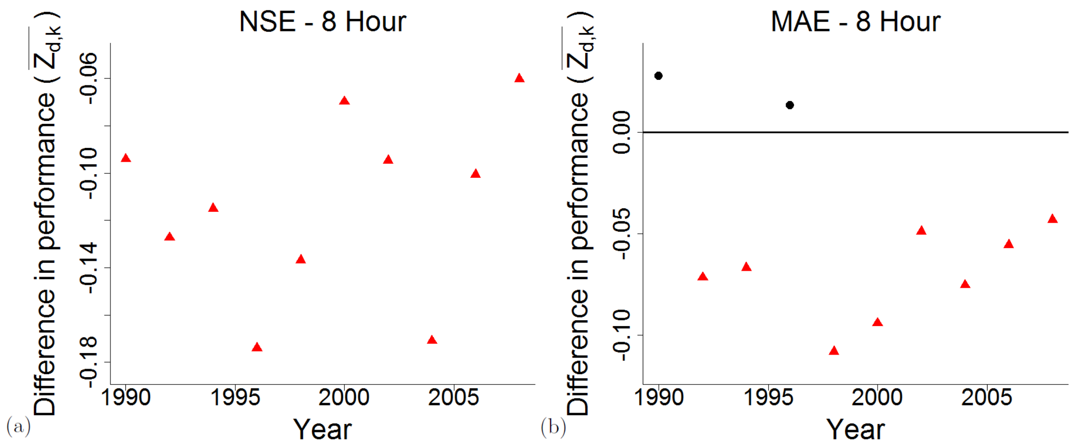


Figure 5-8: Comparison of models using 8-hourly data on River Ray catchment using NSE criteria on the left (a) and MAE on the right (b).

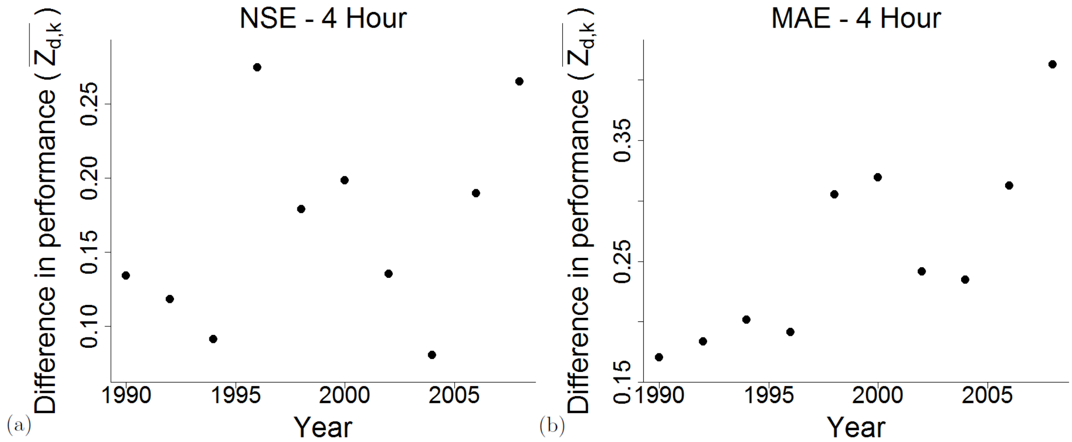


Figure 5-9: Comparison of models using 4-hourly data on River Ray catchment using NSE criteria on the left (a) and MAE on the right (b).

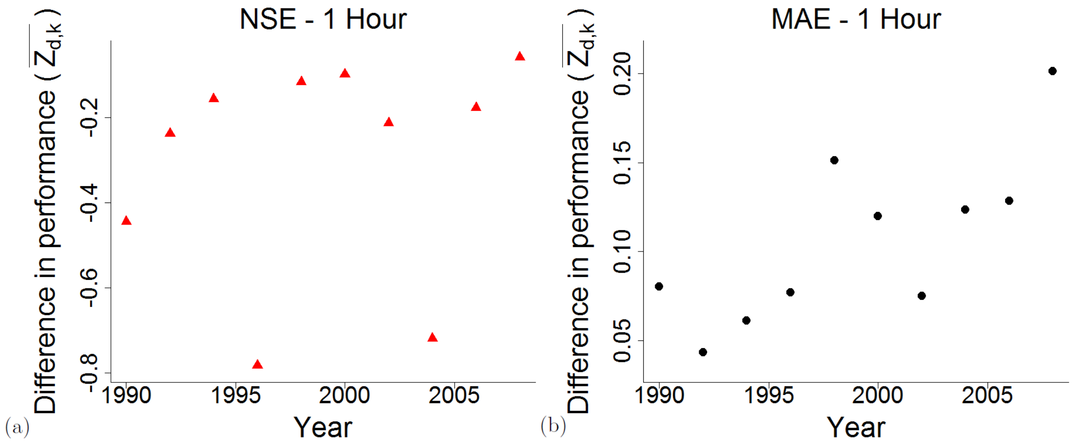


Figure 5-10: Comparison of models using hourly data on River Ray catchment using NSE criteria on the left (a) and MAE on the right (b).

Figures 5-7-5-10 show that for 12-hourly and 4-hourly data URMOD$_1$ outperforms URMOD$_2$ for every two-year validation period, whereas for 8-hourly the reverse is true. The reason for the alternating performance is during the 8 hour time step the difference in performance is very small with differences of between 0.04 and 0.17, this means that performance is very similar and could indicate model performance is the same. But at the 4-hourly time step is when the urban signal is detectable and URMOD$_1$ will perform better. Finally, when looking at hourly data, the performance criteria give contrasting conclusions. The NSE states that URMOD$_2$ has outperformed URMOD$_1$ whereas the reverse is true for the MAE. The reason for this is due to the nature of the performance criteria, since the NSE is a squared measure of performance so larger differences in performance are exaggerated. Whilst the MAE takes a absolute measure of error across the data. This indicates that URMOD$_1$ performs better on average than URMOD$_2$, but has larger differences in performance.

Time step	Average URMOD$_1$ NSE	Average URMOD$_1$ MAE	Average URMOD$_2$ NSE	Average URMOD$_2$ MAE
12-hourly	0.70	0.45	0.36	0.9
8-hourly	0.49	0.6	0.61	0.55
4-hourly	0.68	0.45	0.51	0.71
hourly	0.22	0.63	0.52	0.73

Table 5.1: Average performance criteria for NSE and MAE for both URMOD$_1$ and URMOD$_2$. Larger NSE values show better performance whereas smaller values of MAE show better performance

Table 5.1 shows the average performance criteria for both URMOD$_1$ and URMOD$_2$, for each time step and criteria. As expected 12-hourly and 4-hourly performance criteria were the best performing for URMOD$_1$ whereas a decrease is observed in performance for 8-hourly data. In contrast with URMOD$_2$ in which the worst performing time step was 12-hourly and 4-hourly, and the best performing is 8-hourly. Whilst the worst performance was observed when using hourly data NSE and MAE average for URMOD$_1$, but not for URMOD$_2$.

Time step	p-value	outcome
12-hourly	2.2x10^{-16}	H_0 is rejected such that there is a significant difference in model performance in favour of URMOD$_1$.
8-hourly	3x10^{-13}	H_0 is rejected such that there is a significant difference in model performance in favour of URMOD$_2$.
4-hourly	2.2x10^{-16}	H_0 is rejected such that there is a significant difference in model performance in favour of URMOD$_1$.
hourly	1.43x10^{-07}	H_0 is rejected such that there is a significant difference in model performance in favour of URMOD$_2$.

Table 5.2: Binomial hypothesis test for the NSE performance criteria

Time step	p-value	outcome
12-hourly	2.2×10^{-16}	H_0 is rejected such that there is a significant difference in model performance in favour of $URMOD_1$.
8-hourly	0.002	H_0 is rejected such that there is a significant difference in model performance in favour of $URMOD_2$.
4-hourly	2.2×10^{-16}	H_0 is rejected such that there is a significant difference in model performance in favour of $URMOD_1$.
hourly	2.98×10^{-14}	H_0 is rejected such that there is a significant difference in model performance in favour of $URMOD_1$.

Table 5.3: Binomial hypothesis test for the MAE performance criteria

Table 5.2 and 5.3 show the binomial hypothesis test results for 12-hourly, 8-hourly, 4-hourly and hourly data for both the MAE and NSE results. For every instance the H_0 is rejected such that there is a significant difference in model performance in favour of a particular model. The results obtained as expected and reflect the results presented in Figures 5-7, - 5-10.

5.5.2 Comparison of simulated flow and observed flow from $URMOD_1$ and $URMOD_2$ on the River Ray catchment

This section presents simulated flow from both the $URMOD_1$ and $URMOD_2$ model for selected years. All of the figures presented in this section are from the validation period and no calibration results are presented. The model parameters were chosen based upon the model calibration performance criteria being the largest for NSE and smallest for MAE in order to show the two models "best" performance.

Comparison of simulated flow and observed flow from $URMOD_1$ and $URMOD_2$ for daily and 12-hourly data

Figure 5-11 shows the observed daily rainfall, observed and simulated daily flow from $URMOD_1$ and $URMOD_2$ for the calender year 1997, and observed and simulated 12-hourly flow from $URMOD_1$ and $URMOD_2$ for the year 1997.

The results in Figure 5-11 show that for both daily and 12 hourly aggregated data, $URMOD_1$ generally matches the low flows whilst $URMOD_2$ does not.

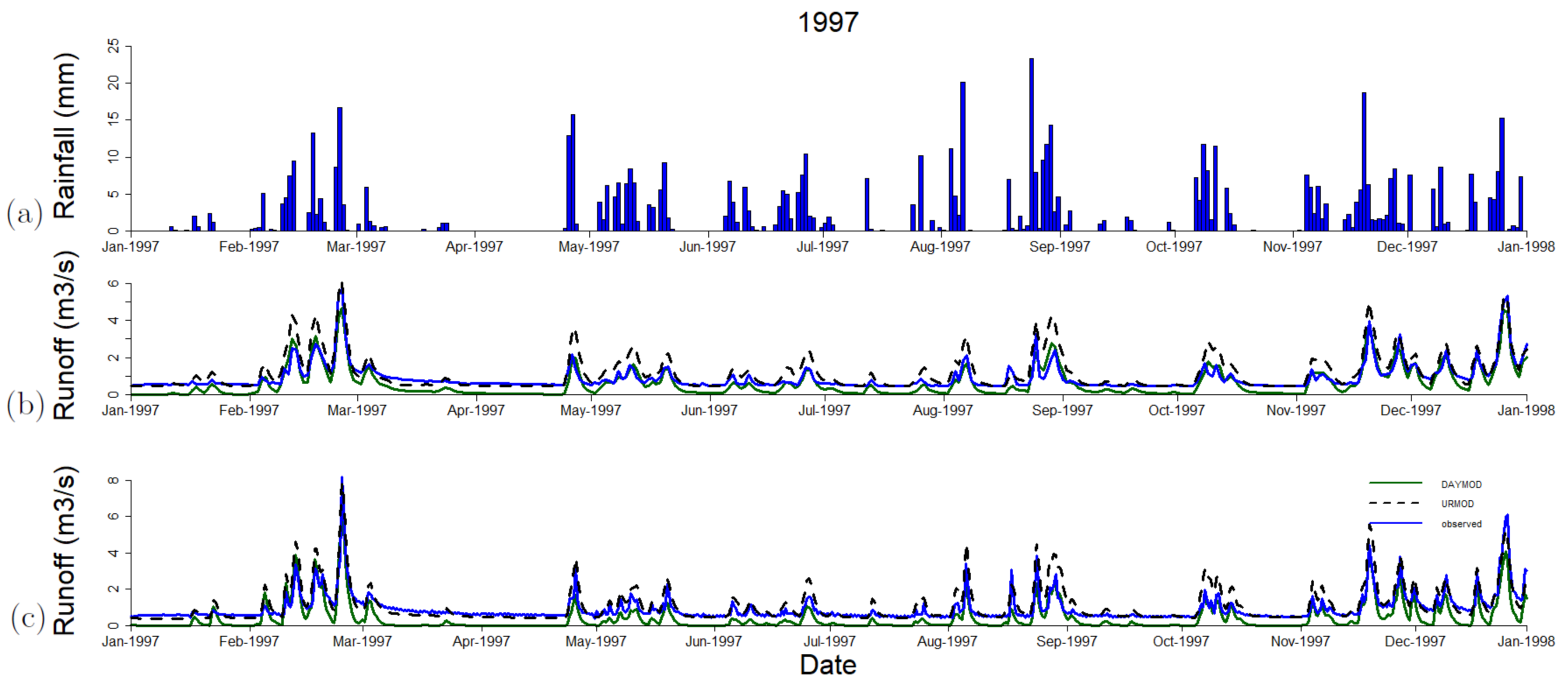


Figure 5-11: The year 1997, top figure (a) observed daily rainfall (mm), middle figure (b) observed daily river flow (blue solid line), $URMOD_1$ simulated daily river flow (black dashed line), $URMOD_2$ estimated daily river flow (green solid line), bottom figure (c) observed 12-hourly river flow (blue solid line), $URMOD_1$ estimated 12-hourly river flow (black dashed line), $URMOD_2$ estimated 12-hourly river flow (green solid line)

When using daily data, stream flow simulated by $URMOD_1$ does over-estimate during most of the peaks, except during the month of December 1997, where as stream flow simulated by $URMOD_2$ both over and under-estimates observed flow with some smaller peaks being estimated very accurately. Using 12 hourly data $URMOD_2$ again consistently under-estimates observed flow and $URMOD_1$ still tends over-estimate observed peaks, but to a lesser degree than observed daily data.

Figure 5-12 shows the observed daily rainfall, observed and simulated daily and 12-hourly flow from $URMOD_1$ and $URMOD_2$ for the calender year 2005. The results in Figure 5-12, similar to Figure 5-11 show $URMOD_1$ matches the low flow for both daily and 12 hourly observed flow, with $URMOD_2$ still unable to simulate low flow. $URMOD_1$ does still over-estimate observed peaks but are not as large as shown in Figure 5-11, and estimates the peaks for the daily simulation a lot more closely than that of $URMOD_2$. The same conclusion can also be drawn for the 12 hour hydrograph. It is noticeable that the peak at April 2005 on Figure 5-12 for daily data is estimated closely by $URMOD_1$ but for 12 hourly does not appear to reach half the observation, indicating either potentially an error in the streamflow data or that when $URMOD_1$ calibrated prioritised calibrating for low flows and smaller peaks.

Comparison of simulated flow and observed flow from $URMOD_1$ and $URMOD_2$ for 8-hourly, 4-hourly and hourly data

For the 8-hourly, 4-hourly and hourly data a section from each year is selected for the hydrograph analysis. This is used for analysis of performance of the model for periods of less rainfall and periods of more constant rainfall. The periods selected are April to July 1997, and October to December 2005. These periods were selected because April to July had on average 45-55 mm of rainfall for the river Ray catchment, whereas the river Cut catchment had 40-50 mm of rainfall. This is opposed to October to December which had on average 70-75 mm for the river Ray catchment with the river Cut having 55-65 mm of rainfall. One potential hindrance to performance is shown in Figure 5-13 and 5-14 is the oscillations in all of the observed flow hydrographs. This could indicate waste-water treatment plants discharging small amounts into the river. This could impact model calibration and provide reasoning for poor performance during the

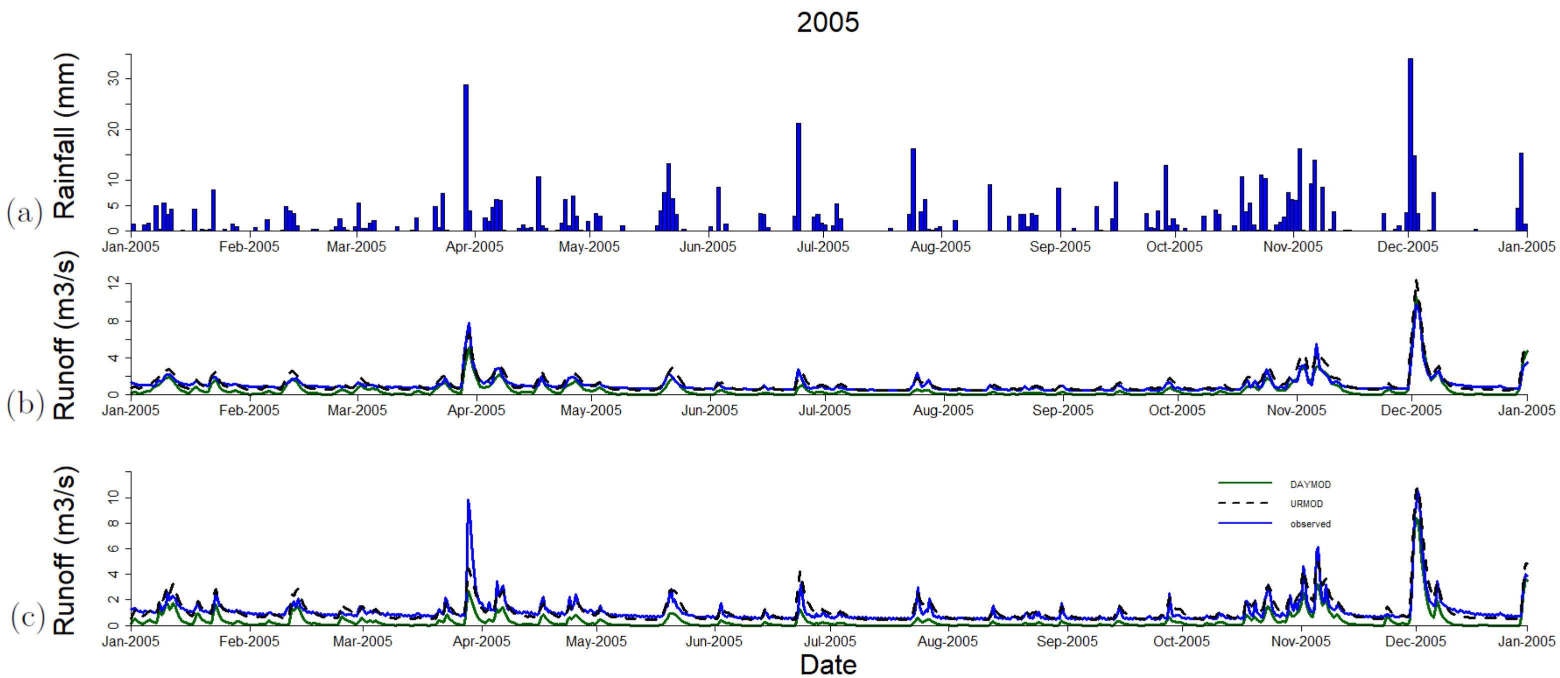


Figure 5-12: The year 2005, top figure (a) observed daily rainfall (mm), middle figure (b) observed daily river flow (blue solid line), $URMOD_1$ simulated daily river flow (black dashed line), $URMOD_2$ estimated daily river flow (green solid line), bottom figure (c) observed 12-hourly river flow (blue solid line), $URMOD_1$ estimated 12-hourly river flow (black dashed line), $URMOD_2$ estimated 12-hourly river flow (green solid line)

8-hourly time step in the previous section.

Figure 5-13 shows the months April to July 1997 and includes observed hourly rainfall, 8-hourly observed and simulated flow from $URMOD_1$ and $URMOD_2$, 4-hourly observed and simulated flow from $URMOD_1$ and $URMOD_2$ and hourly observed and simulated flow from $URMOD_1$ and $URMOD_2$. The results in Figure 5-13 show that $URMOD_1$ is capable of matching the baseflow of the observed river flow, however $URMOD_2$ is unable to do this with flow tending towards zero after rain events. For the 8-hour simulations it can be seen that $URMOD_1$ does over-estimate slightly in the second half of the time period, but as the time step is decreased this over-estimation is reduced. During the first half of the simulation $URMOD_1$ matches the observed flow with only very slight over-estimation. This is opposed to $URMOD_2$ which is inconsistent, with the simulated runoff for 8-hourly simulations matching the observed flow after June. But for 4-hourly under-estimates for the entire time period, whilst matching observed flow for the hourly time step. Figure 5-14 show observed hourly rainfall, 8-hourly observed and simulated flow from $URMOD_1$ and $URMOD_2$, 4-hourly observed and simulated flow from $URMOD_1$ and $URMOD_2$ and hourly observed and simulated flow from $URMOD_1$ and $URMOD_2$; for the period Oct-Dec (2005). The results in Figure 5-14, unlike Figure 5-13, shows that $URMOD_2$ does estimate flow with only slight under-estimation in during the end of October and beginning of November. $URMOD_2$ is unable to estimate the baseflow for 4-hourly and hourly data. $URMOD_1$ does again estimate the baseflow consistently except for 8-hours with a slight tending for over-estimation. During the peak in the beginning of November $URMOD_1$ under-estimates observed flow, but this is consistent for each time step, whilst the events during October and November matches the observed flow. Whereas for the rainfall event at the beginning of December 2005, both models estimate the peak then the decline whereas the observed data has a gradual rise to the peak, indicating that both models lag time is potentially inaccurate for larger rainfall events.

The analysis for $URMOD_1$ from the river Ray catchment shows that $URMOD_1$ does over and under simulate in some cases but also manages to maintain peak shapes, alongside this $URMOD_1$ does consistently match the baseflow. Whereas $URMOD_2$ consistently under or over simulates alot, can not maintain consistent baseflow and in cases is unable to match the lag times.

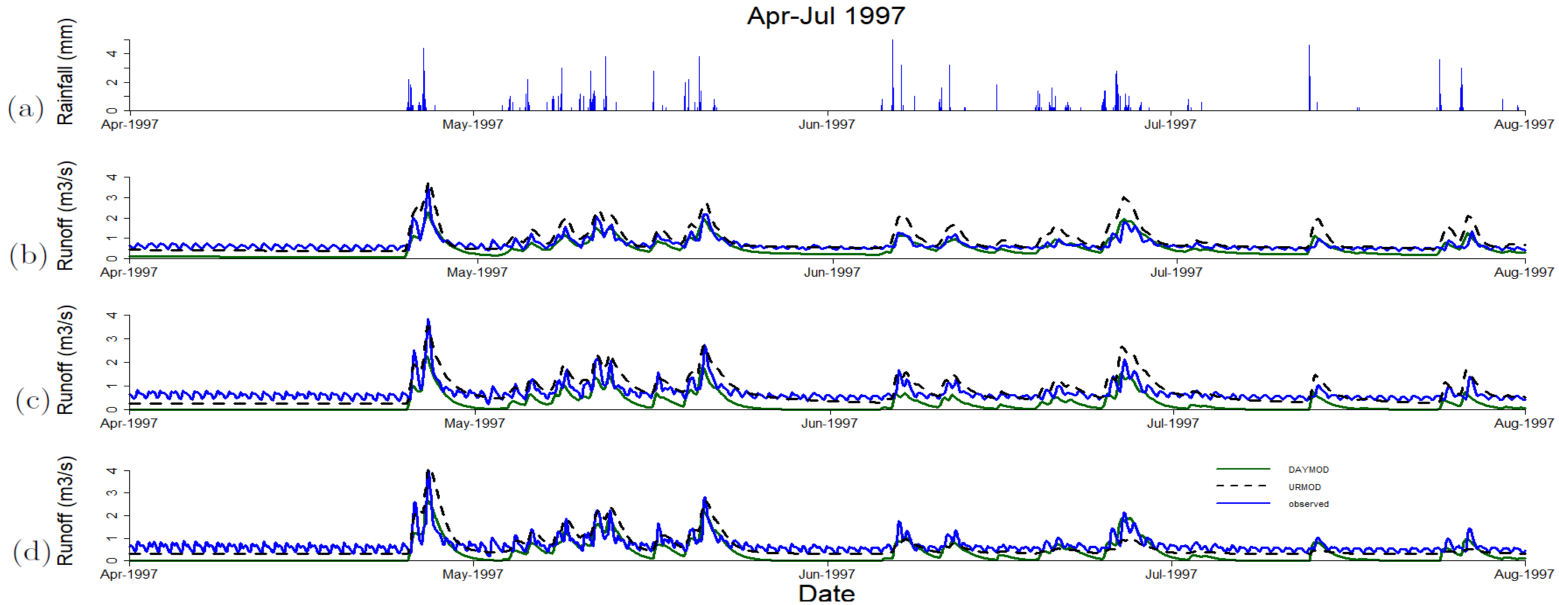


Figure 5-13: The months April to July 1997, top figure (a) observed hourly rainfall (mm), the second figure (b) observed 8-hourly river flow (blue solid line), $URMOD_1$ simulated 8-hourly river flow (black dashed line), $URMOD_2$ estimated 8-hourly river flow (green solid line), third figure (c) observed 4-hourly river flow (blue solid line), $URMOD_1$ estimated 4-hourly river flow (black dashed line), $URMOD_2$ estimated 4-hourly river flow (green solid line), final figure (d) observed hourly river flow (blue solid line), $URMOD_1$ estimated hourly river flow (black dashed line), $URMOD_2$ estimated hourly river flow (green solid line)

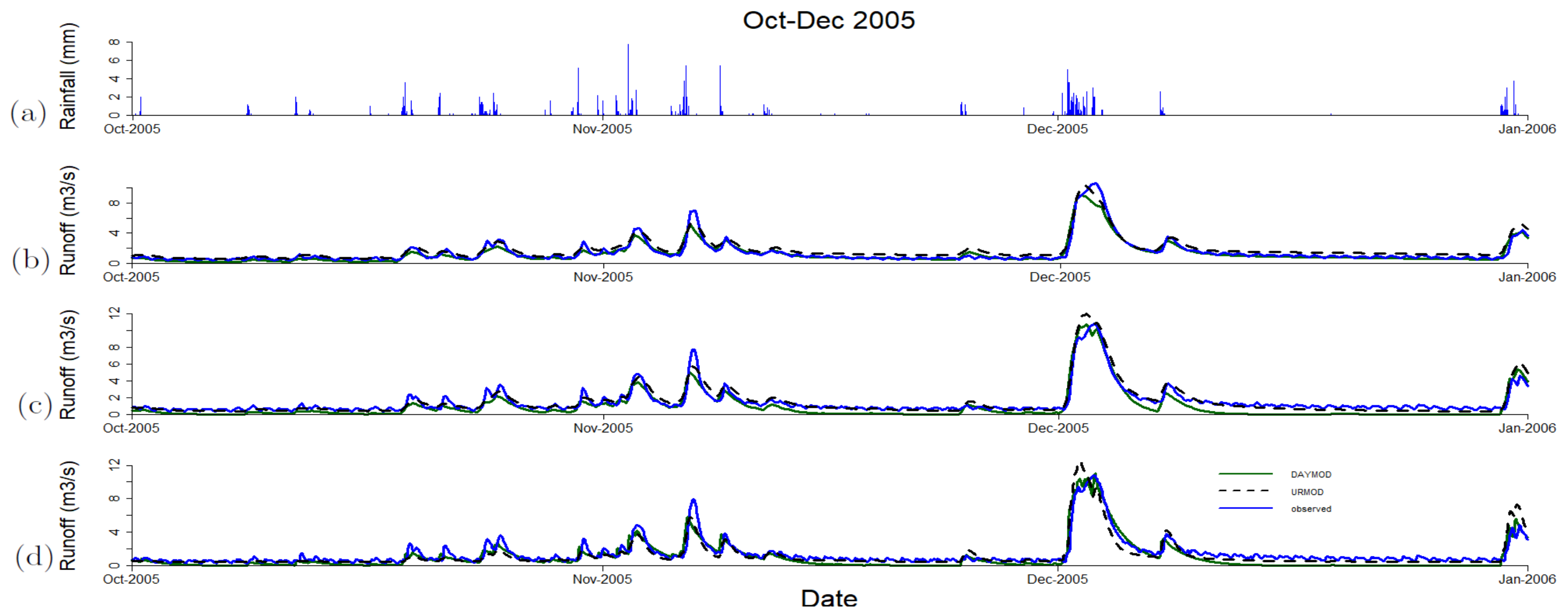


Figure 5-14: The months October to December 1997, top figure (a) observed hourly rainfall (mm), the second figure (b) observed 8-hourly river flow (blue solid line), $URMOD_1$ simulated 8-hourly river flow (black dashed line), $URMOD_2$ estimated 8-hourly river flow (green solid line), third figure (c) observed 4-hourly river flow (blue solid line), $URMOD_1$ estimated 4-hourly river flow (black dashed line), $URMOD_2$ estimated 4-hourly river flow (green solid line), final figure (d) observed hourly river flow (blue solid line), $URMOD_1$ estimated hourly river flow (black dashed line), $URMOD_2$ estimated hourly river flow (green solid line)

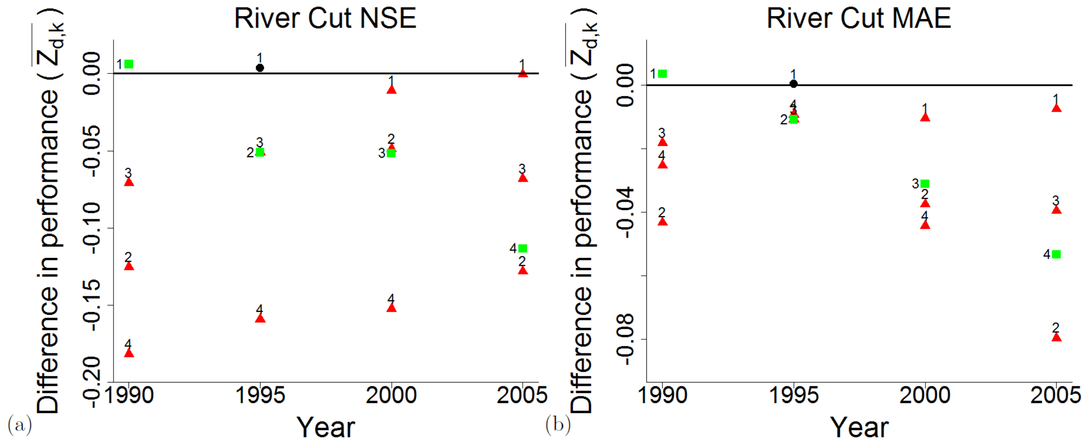


Figure 5-15: Comparison of models on river Cut using NSE criteria on the left (a) and MAE on the right (b).

5.5.3 Assessing performance of URMOD$_1$ and URMOD$_2$ on the River Cut catchment

This section explores the difference in performance of the URMOD$_1$ and URMOD$_2$ models on the river Cut catchment. Comparing performance at the daily time scale first then exploring sub-daily model performance.

Daily comparison of URMOD$_1$ and URMOD$_2$

Figure 5-15 shows the difference in results between URMOD$_1$ and URMOD$_2$ for the River Cut catchment using daily data when using the NSE and MAE performance criteria. Similar to Section 5.5.1 the left hand figure (a) is the difference in NSE, reported as $Z_{1,m,k} - Z_{2,m,k}$ and the right hand figure (b) is the difference in MAE, it is reported as $Z_{2,m,k} - Z_{1,m,k}$. Positive values indicate that URMOD$_1$ performs better than URMOD$_2$, indicated on Figure 5-15 as black circles. In contrast, negative values indicate that URMOD$_2$ performs better than URMOD$_1$. Performance criteria were obtained for the four calibration periods for both models, the difference of the calibration results are presented on Figure 5-15 as green squares. Lastly the numbers next to the points indicate which calibration period they are from, hence 1 indicates the calibration square's parameters were used for the rest of the points with a 1 next to them.

Figure 5-15 show that URMOD$_2$ has outperformed URMOD$_1$ for all but 1 occasion. The average NSE performance for $URMOD_1$ was 0.681 whereas $URMOD_2$ was 0.757, both NSE values would still be classified as good. Whereas the average values for MAE was 0.181 for $URMOD_1$ and for $URMOD_2$ was 0.156, both of these MAE values again would be classified as being good. One potential reason for $URMOD_1$ having a lower performance criteria than $URMOD_2$ is that the rainfall and runoff data showed that the river hydrology of the river Cut is different than that of the river Ray. Such that the from the years 1990-1999, two distinct different periods of flow is observed, months April to October the peaks are flashy and base flow is more sustained, where as the months November to March, flow is a lot more lagged and rain events appear to be sustained, this can be seen in Figure 5-16.

Whereas for the years 2000-2009 this effect appears less frequently. When analysing the plotted river flow data for the river cut for months April to October will be named dry season, where as months November to March will be called wet season. Figure 5-16 shows the observed daily rainfall, observed and simulated daily and 12-hourly flow from $URMOD_1$ and $URMOD_2$ for the calender year 1997.

The results in Figure 5-16 show that for the dry season $URMOD_1$ estimates the observed peak flow and can also match the base flow with some under-estimation, whereas for $URMOD_2$ it can not estimate the base flow and under-estimates frequently when compared with the peaks. This conclusion can be drawn for both daily and 12-hourly time step. Whereas for the wet season, neither $URMOD_1$ and $URMOD_2$ can estimate the base flow, with under-estimation for both time steps. For the daily data $URMOD_2$ matches the peaks but it is a lot more flashy than that of $URMOD_1$, however during 12-hourly simulation $URMOD_2$ does over-estimate. This is opposite to $URMOD_1$ which during daily time steps only has slight over-estimation, but during 12-hour has slight under estimation.

Figure 5-17 shows the observed daily rainfall, observed and simulated daily and 12-hourly flow from $URMOD_1$ and $URMOD_2$ for the calender year 2005. Figure 5-17 indicates that similar to the results presented in Figure 5-16 neither $URMOD_1$ or $URMOD_2$ is capable of consistently estimating the base flow for both observed daily and 12 hourly data. $URMOD_1$ under and over estimates for both the daily/ 12-hourly data, whilst $URMOD_2$ only under-estimated. Both $URMOD_1$ and $URMOD_2$ are capable of estimating the peaks however both

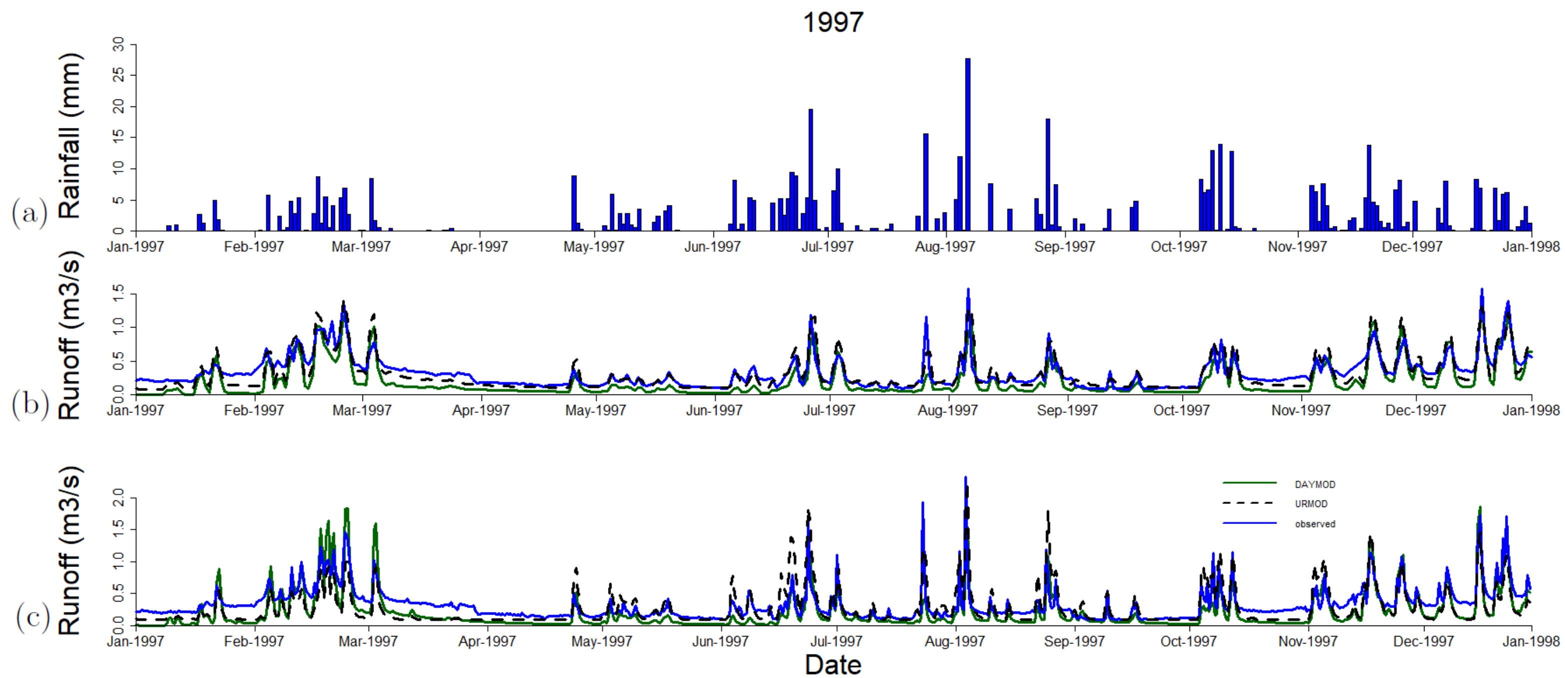


Figure 5-16: The year 1997, top figure (a) observed daily rainfall (mm), middle figure (b) observed daily river flow (blue solid line), $URMOD_1$ simulated daily river flow (black dashed line), $URMOD_2$ estimated daily river flow (green solid line), bottom figure (c) observed 12-hourly river flow (blue solid line), $URMOD_1$ estimated 12-hourly river flow (black dashed line), $URMOD_2$ estimated 12-hourly river flow (green solid line).

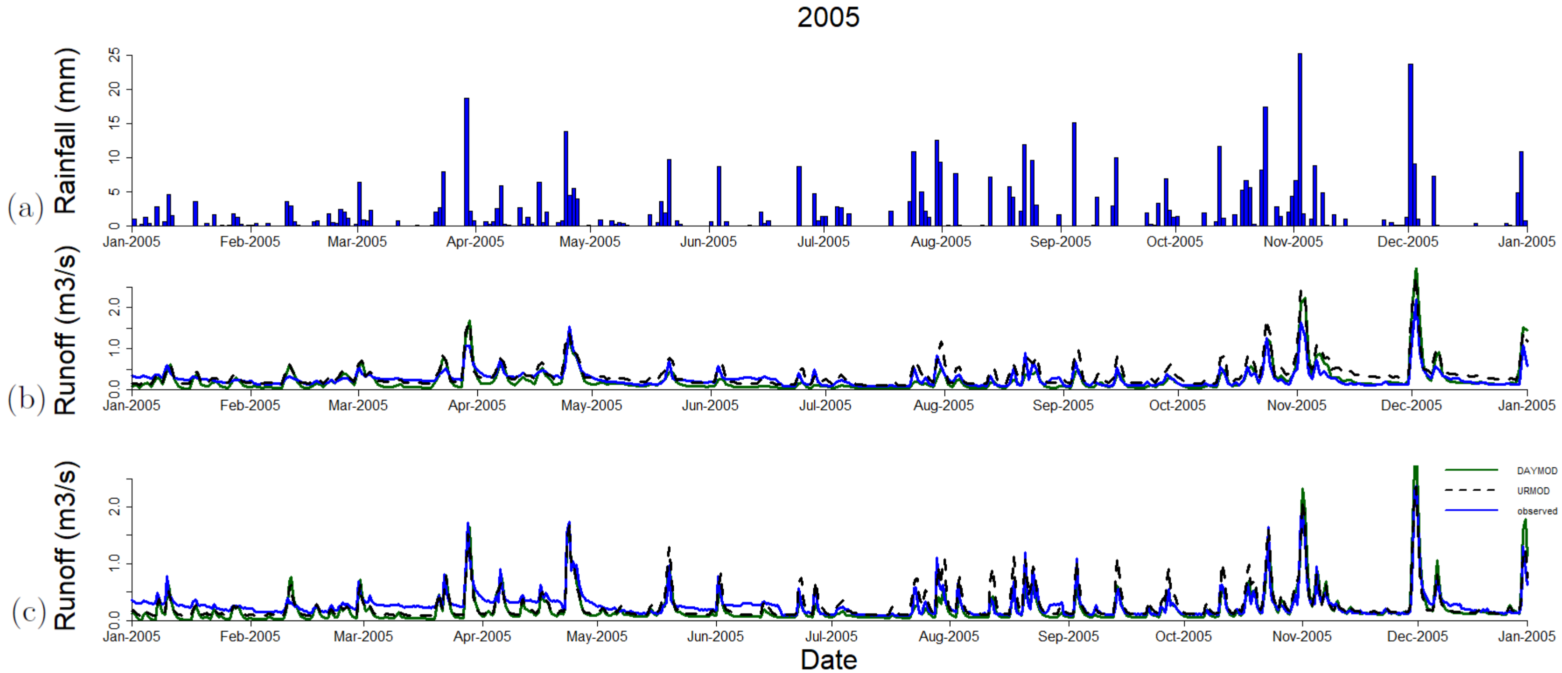


Figure 5-17: The year 2005, top figure (a) observed daily rainfall (mm), middle figure (b) observed daily river flow (blue solid line), $URMOD_1$ simulated daily river flow (black dashed line), $URMOD_2$ estimated daily river flow (green solid line), bottom figure (c) observed 12-hourly river flow (blue solid line), $URMOD_1$ estimated 12-hourly river flow (black dashed line), $URMOD_2$ estimated 12-hourly river flow (green solid line).

$URMOD_1$ and $URMOD_2$ over and under-estimated consistently the observed data for both daily and 12 hourly data.

Subdaily comparison of $URMOD_1$ and $URMOD_2$ on the river Cut catchment

Table 5.4 shows the average performance criteria for both $URMOD_1$ and $URMOD_2$ for the performance criteria NSE and MAE for each time step.

Time step	Average $URMOD_1$ NSE	Average $URMOD_1$ MAE	Average $URMOD_2$ NSE	Average $URMOD_2$ MAE
12-hourly	0.55	0.2	0.71	0.18
8-hourly	0.25	0.49	0.31	0.26
4-hourly	0.66	0.18	0.72	0.31
hourly	0.11	0.29	0.26	0.5

Table 5.4: Average performance criteria for NSE and MAE for both $URMOD_1$ and $URMOD_2$. Larger NSE values show better performance whereas smaller values of MAE show better performance.

Similar to the river Ray, the $URMOD_1$ models best performing time step is 12-hourly and 4-hourly, with a drop in performance for both 8-hourly and hourly. Unlike the river Ray catchment the performance of $URMOD_2$ is better than that of $URMOD_1$ for each time step with the NSE, but for the MAE only has a smaller value on two occasions. With 12-hourly only being a difference of 0.02. One reason for such a difference in performance is outlined in Section 5.5.3 such that the river flow has two distinct patterns in which both models can't simulate and so taking a lumped value of two years produces inaccurate performance criteria. The binomial hypothesis test results are presented in Table 5.5 and 5.6.

For one instance the H_0 can not be rejected such that there is no significant difference in model performance using the NSE for 8-hourly data. The rest of the binomial hypothesis tests showed that H_0 can be rejected such that there is a difference in performance in favour of either $URMOD_1$ or $URMOD_2$. The results obtained form the binomial hypothesis are expected due to the average NSE and MAE results presented in Table 5.5 and 5.6

Time step	p-value	outcome
12-hourly	$5.202\text{x}10^{-13}$	H_0 is rejected such that there is a significant difference in model performance in favour of URMOD_2.
8-hourly	0.07255	H_0 is accepted such that there is a no significant difference in model performance.
4-hourly	$4.386\text{x}10^{-10}$	H_0 is rejected such that there is a significant difference in model performance in favour of URMOD_2.
hourly	$4.386\text{x}10^{-10}$	H_0 is rejected such that there is a significant difference in model performance in favour of URMOD_2.

Table 5.5: Binomial hypothesis test for the NSE performance criteria

Time step	p-value	outcome
12-hourly	$3.795\text{x}10^{-06}$	H_0 is rejected such that there is a significant difference in model performance in favour of URMOD_2.
8-hourly	$2.2\text{x}10^{-16}$	H_0 is rejected such that there is a significant difference in model performance in favour of URMOD_2.
4-hourly	$2.2\text{x}10^{-16}$	H_0 is rejected such that there is a significant difference in model performance in favour of URMOD_1.
hourly	$2.2\text{x}10^{-16}$	H_0 is rejected such that there is a significant difference in model performance in favour of URMOD_1.

Table 5.6: Binomial hypothesis test for the MAE performance criteria

Comparison of simulated flow and observed flow from URMOD_1 and URMOD_2 for 8-hourly, 4-hourly and hourly data

Similar to Section 5.5.2 the observed and simulated flow for 8-hourly, 4-hourly and hourly data is presented as sections from a year as opposed to the full year. Figure 5-18 is the calender year 1997 from the month of April to July. Whilst Figure 5-19 is the months October to December for the calender year 2005. Both Figures show observed hourly rainfall, 8-hourly observed and simulated flow from URMOD_1 and URMOD_2, 4-hourly observed and simulated flow from URMOD_1 and URMOD_2 and hourly observed and simulated flow from URMOD_1 and URMOD_2.

The results from Figure 5-18 show that both URMOD_1 and URMOD_2 are incapable of matching the baseflow of the observed river flow whereas the results in Figure 5-19 show that both models can simulate it with only slight under and over estimation. From both Figures the simulated riverflow from URMOD_2 for the

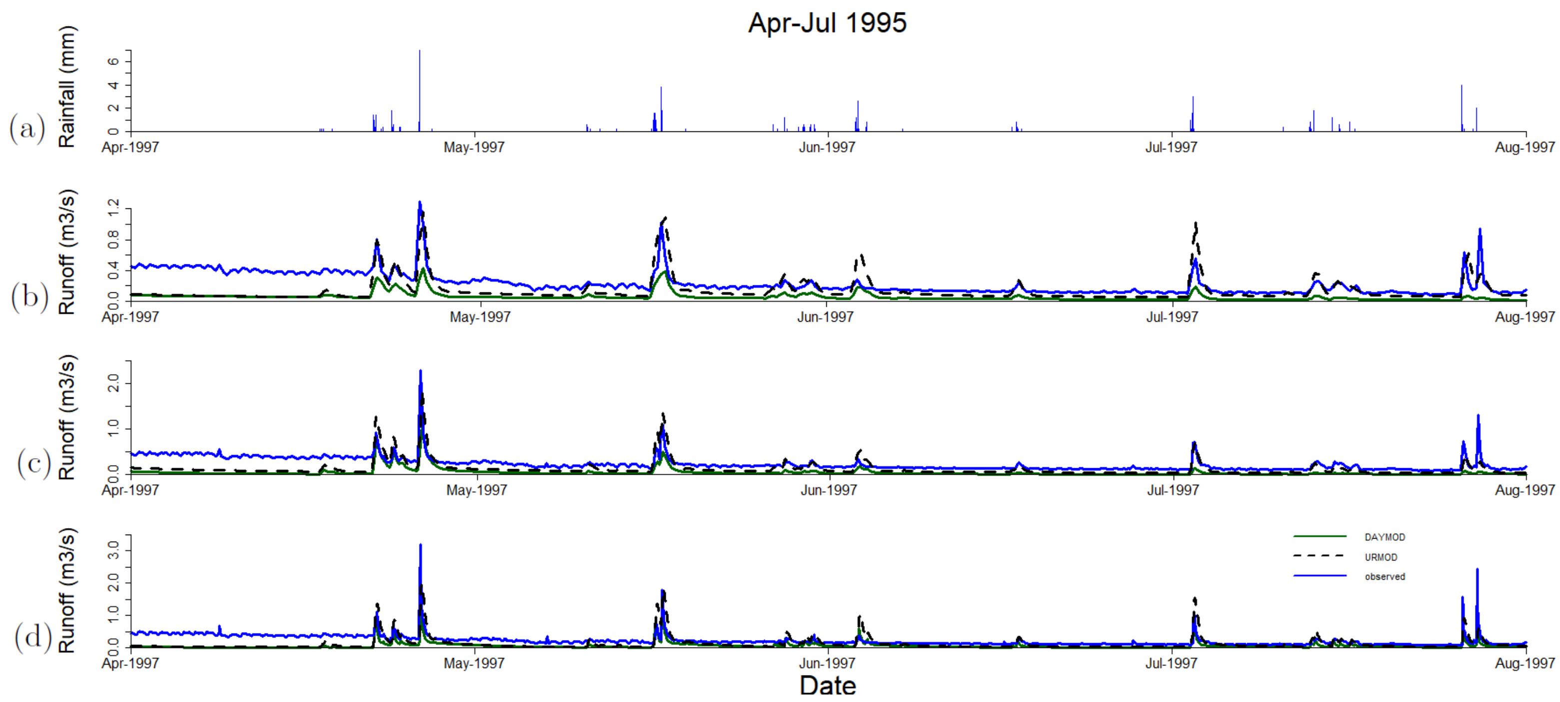


Figure 5-18: The months April to July 1997, top figure (a) observed hourly rainfall (mm), the second figure (b) observed 8-hourly river flow (blue solid line), $URMOD_1$ simulated 8-hourly river flow (black dashed line), $URMOD_2$ estimated 8-hourly river flow (green solid line), third figure (c) observed 4-hourly river flow (blue solid line), $URMOD_1$ estimated 4-hourly river flow (black dashed line), $URMOD_2$ estimated 4-hourly river flow (green solid line), final figure (d) observed hourly river flow (blue solid line), $URMOD_1$ estimated hourly river flow (black dashed line), $URMOD_2$ estimated hourly river flow (green solid line).

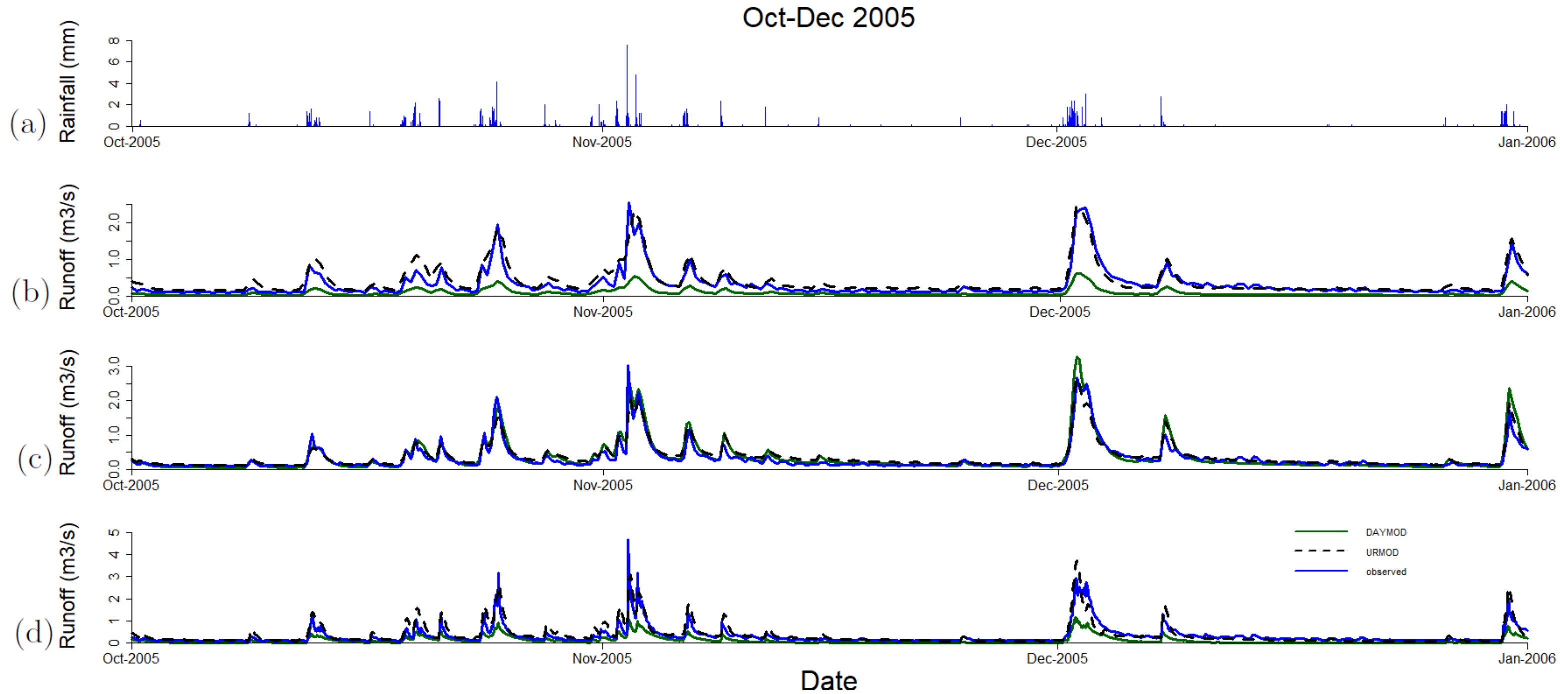


Figure 5-19: The months October to December 1997, top figure (a) observed hourly rainfall (mm), the second figure (b) observed 8-hourly river flow (blue solid line), $URMOD_1$ simulated 8-hourly river flow (black dashed line), $URMOD_2$ estimated 8-hourly river flow (green solid line), third figure (c) observed 4-hourly river flow (blue solid line), $URMOD_1$ estimated 4-hourly river flow (black dashed line), $URMOD_2$ estimated 4-hourly river flow (green solid line), final figure (d) observed hourly river flow (blue solid line), $URMOD_1$ estimated hourly river flow (black dashed line), $URMOD_2$ estimated hourly river flow (green solid line).

12-hourly time step shows that the model consistently under-estimates the peaks. Whilst URMOD_1 is capable of simulating the peaks with slight over-estimation, which is in conflict with the results presented in Table 5.4 that URMOD_2 performs better on 12-hourly flow. For the 4-hourly simulations for the year 2005 both models consistently match the observed riverflow, with only slight over-estimation on the December 2005 peaks. Whereas for the year 1997 URMOD_1 only over-estimates slightly expect for the peak in late July. Conversely URMOD_2 under-simulates the observed flow. For the hourly simulations similar to the 12-hourly simulations URMOD_2 consistently under-estimates for both the years 2005 and 1997. Whilst URMOD_1 for both years 2005 and 1997 is capable in simulating the observed flow with only slight over-estimation.

5.6 Discussion

The results presented in this paper are used to address the research question of whether a simple urban framework implemented onto an existing rural rainfall-runoff model can significantly improve performance in urban catchments. Results indicate that whilst the urban model URMOD_1 is capable of producing accurate simulations a number of issues need to be considered.

Firstly, when comparing the performance of the urban and rural models for daily data, the results for the river Ray catchment showed that the urban model outperformed the rural model. However, conflicting conclusions are drawn for the river Cut catchment. Whilst the performance criteria indicated that the rural model performed better, when analysing the hydrographs resulted showed that this conclusion might be misleading. To solve this, rather than having one lumped performance criteria for the entire period, splitting each year into sections and applying a performance criteria for each. This will allow for differences between more and less periods of rainfall to be analysed better.

The size of the catchments may have impacted the performance criteria results as outlined by Salvadore et al. (2015) that urban flow routing is unlikely to be effective on the daily scale for smaller catchments. Whilst both catchments are similar in size, the daily results presented support Salvadore et al. (2015) such that for the smaller catchment (Cut 50.2km^2) the results were worse than the larger catchment (Ray 81.6km^2), indicating that for the smaller catchment the inclusion of urban routing potentially was ineffective.

Similarly Mitchell et al. (2001) argues that excess rainfall has left urban catchments in a matter of hours. This conclusion is supported by the results found in this study as the $URMOD_1$ model best performing criteria occurred when using the 4-hourly time step and from the hydrographs presented the level of over simulation decreased as the time step got smaller indicating that the urban routing performed better at this time scale. One important observation is the difference in performance when considering the hydrographs for 1997 and 2005. For the river Ray catchment results from 1997 showed URMOD over-estimated consistently whilst the same tendency was not observed for 2005, where only slight under or over estimations occurred. One potential reason for this is due to two levels of urbanisation being used ($URBEXT_{1990}$ and $URBEXT_{2000}$) since over-estimation during 1997 could indicate that $URBEXT_{1990}$ was larger than the actual value. However this conclusion can not be drawn for the river Cut catchment as for both years over and under estimations occurred.

One large problem with $URMOD_1$ is the calibration during finer time stepped data (hourly/ 4-hourly). This is because of long periods of time during which no rainfall is present and only constant flow which can influence parameter estimation. In order to account for this during the calibration of the model, only peak flows and rainfall events are considered by the objective function. As mentioned in the beginning of this section the flow for the river Cut catchment differs depending on the time of year. Such that during the winter months a larger base flow is observed, whilst during the summer months less base flow is present. In order to account for this a multi-parameter strategy could be employed, such that during winter months one parameter set is used whereas during the summer months an alternative parameter set can be used.

5.7 Conclusion

This study used a combination of performance criteria, hydrographs and binomial hypothesis tests in order to determine if a simple urban framework is effective at modelling urban catchments when using sub daily data. The results showed that for the larger catchment (river Ray) the urban model outperformed the rural version of the model and was capable of simulating both the peaks and the baseflow.

But for the smaller catchment (river Cut) results indicated that the rural model performed better. This paper presents a simple urban framework in order to account for two urban processes (runoff and pipe routing) and whilst conflicting conclusions are drawn for two catchments the URMOD model is designed to take simple assumptions of urban processes and is a starting point for further urban processes to be implemented into the model with further research.

Acknowledgments

The authors would like to thank the National River Flow Archive (NRFA) for providing access to hydrological data. The NERC funded POLLCURB project for providing access to the hydrological and land-use data used in this study (NE/K002317/1).

5.A URMOD and DAYMOD input, output and parameters

Parameter	Description	Unit
S	Average Soil moisture capacity	mm
F	Field capacity	mm
R	Rooting Depth	mm
k	Drainage coefficient	-
Φ	Proportion of runoff split	-
B_L	Baseflow lag	days
S_L	Channel lag	days
U_L	Urban lag	days
γ	Scaling term	-

Table 5.7: URMOD model parameters, description and units

Input	abbreviation	Unit
Observed rainfall	i	mm
Observed riverflow	q_{obs}	m^3/s
Potential evaporation	E_p	mm
Ouput	abbreviation	Unit
Simulated riverflow	q_{sim}	m^3/s
Simulated baseflow	b_t	m^3/s
Average soil moisture	-	mm
Actual evaporation	-	mm

Table 5.8: URMOD inputs and outputs

5.2 Bibliography

Bayliss, A., Black, K., Fava-Verde, A., and Kjeldsen, T. (2006). URBEXT2000 - a new FEH catchment descriptor. calculation, dissemination and application. Technical Report R&D FD1919/TR, Department for Environment Food and Rural Affairs, CEH wallingford.

Butts, M. B., Payne, J. T., Kristensen, M., and Madsen, H. (2004). An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *Journal of Hydrology*, 298(1):242–266.

Duan, Q., Gupta, V. K., and Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Applications*, 76(3):501–521.

Fletcher, T., Andrieu, H., and Hamel, P. (2013). Understanding, management and modelling of urban hydrology and its consequences for receiving waters: A state of the art. *Advances in Water Resources*, 51:261–279.

Jacobson, C. R. (2011). Identification and quantification of the hydrological impacts of imperviousness in urban catchments: A review. *Journal of Environmental Management*, 92(6):1438–1448.

Jakeman, A. and Hornberger, G. (1993). How much complexity is warranted in a rainfall-runoff model? *Water Resources Research*, 29(8):2637–2649.

Jones, D. A. and Kay, A. L. (2007). Uncertainty analysis for estimating flood frequencies for ungauged catchments using rainfall-runoff models. *Advances in Water Resources*, 30(5):1190–1204.

Jones, J. A., Swanson, F. J., Wemple, B. C., and Snyder, K. U. (2000). Effects of roads on hydrology, geomorphology, and disturbance patches in stream networks. *Conservation Biology*, 14(1):76–85.

Keller, V., Tanguy, M., Prosdocimi, I., Terry, J., Hitt, O., Cole, S., Fry, M., Morris, D., and Dixon, H. (2015). CEH-GEAR: 1 km resolution daily and monthly areal rainfall estimates for the UK for hydrological and other applications. *Earth System Science Data*, 7(1):143–155.

Kjeldsen, T., Stewart, E., Packman, J., Folwell, S., and Bayliss, A. (2005). Revitalisation of the FSR/FEH rainfall-runoff method. Technical report, Defra R&D Technical Report FD1913/TR, CEH Wallingford.

Klemeš, V. (1986). Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, 31(1):13–24.

Li, H., Xu, C.-Y., and Beldring, S. (2015). How much can we gain with increasing model complexity with the same model concepts? *Journal of Hydrology*, 527:858–871.

Loague, K. M. and Freeze, R. A. (1985). A comparison of rainfall-runoff modeling techniques on small upland catchments. *Water Resources Research*, 21(2):229–248.

Miller, J. D., Kim, H., Kjeldsen, T. R., Packman, J., Grebby, S., and Dearden, R. (2014). Assessing the impact of urbanization on storm runoff in a peri-urban catchment using historical change in impervious cover. *Journal of Hydrology*, 515:59–70.

Mitchell, V., Mein, R. G., and McMahon, T. A. (2001). Modelling the urban water cycle. *Environmental Modelling & Software*, 16(7):615–629.

Moore, R. (1985). The probability-distributed principle and runoff production at point and basin scales. *Hydrological Sciences Journal*, 30(2):273–297.

Nash, J. E. and Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I-A discussion of principles. *Journal of Hydrology*, 10(3):282–290.

of Hydrology, I. (1999). *Flood estimation handbook*, volume 4. Institute of Hydrology Wallingford.

Orth, R., Staudinger, M., Seneviratne, S. I., Seibert, J., and Zappa, M. (2015). Does model performance improve with complexity? a case study with three hydrological models. *Journal of Hydrology*, 523:147–159.

Packman, J. (2004). Impact of antecedent rainfall longer than 5 days on base-flow and percentage runoff. Technical Report Environment agency R & D report W6-080, Centre for Ecology and Hydrology, CEH Wallingford.

Perrin, C., Michel, C., and Andréassian, V. (2001). Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. *Journal of Hydrology*, 242(3):275–301.

Putro, B., Kjeldsen, T., Hutchins, M. G., and Miller, J. (2016). An empirical investigation of climate and land-use effects on water quantity and quality in two urbanising catchments in the southern united kingdom. *Science of the Total Environment*, 548:164–172.

Redfern, T. W., Macdonald, N., Kjeldsen, T. R., Miller, J. D., and Reynard, N. (2016). Current understanding of hydrological processes on common urban surfaces. *Progress in Physical Geography*, 40(5):699–713.

Robinson, E.L.Blyth, E. D.-P. E. J. A. (2016). Climate hydrology and ecology research support system potential evapotranspiration dataset for great britain (1961-2015). data retrieved from NERC Environmental Information Data Centre, `https://doi.org/10.5285/8baf805d-39ce-4dac-b224-c926ada353b7`.

Rose, S. and Peters, N. E. (2001). Effects of urbanization on streamflow in the Atlanta area (Georgia, USA): a comparative hydrological approach. *Hydrological Processes*, 15(8):1441–1457.

Salvadore, E., Bronders, J., and Batelaan, O. (2015). Hydrological modelling of urbanized catchments: a review and future directions. *Journal of Hydrology*, 529:62–81.

Shaw, E. (1994). *Hydrology in practice.* (3rd ed.), Chapman and Hall, London.

The previous chapter attempted to answer research question 1 by comparing the full URMOD model and the nested rural model (DAYMOD) on two catchments using sub-daily data. Results show that for the river Ray catchment the URMOD model outperformed the rural model whilst for the river Cut conflicting conclusions are drawn. In order to test the full potential of the URMOD model and fully explore **Research question 1:** Can rainfall-runoff modelling in urban catchments be improved via implementation of a conceptual urban framework onto an already existing rural based model ? A number of different hydrological experiments need to be undertaken. All of which will be applied in the next chapter, these are model performance on a catchment with extreme levels of rainfall and event based modelling. Chapter 6 will explore these alongside answering **Research question 2:** What is the loss of information when moving from lumped to spatially distributed rainfall-runoff models on a varying time scale for urban catchments ? This will explore the performance of the URMOD model by comparing the performance with a semi-distributed model (CAT). The binomial hypothesis test and jackknife calibration/validation method developed in Chapter 4 will also be used in the next chapter to test model performance and to calibrate URMOD. The table below outlines the contribution of the thesis author and the co-authors of the coming paper chapter.

This declaration concerns the article entitled:									
A comparison between lumped and semi-distributed models using sub-daily data on the Gyeongancheon and Rodbourne catchments.									
Publication status (tick one)									
Draft	✓	Submitted		Review		Accepted		Published	
Publication details (reference)	James Fidal, Hyeonjun Kim, Thomas Kjeldsen & Cheolhee Jang. Draft format to be submitted to Journal of the Korean Society of Water and Wastewater (KSWW)								
Candidate's contribution to the paper (detailed, and also given as a percentage).	The author of this thesis contributed to the entire paper, formulating the ideas, defining the methodology used, application of the models and preparation/ writing of the manuscript. Data collection was not handled as part of this paper and was provided by two of the other authors (Hyeonjun Kim and Cheolhee Jang). Each author's exact contribution is detailed below Formulation of ideas: 70% James, 10% Thomas, 10% Cheolhee, 10% Hyeonjun Model application: 80% James 20% Cheolhee Data acquisition: 100% Jang and Hyeonjun Analyse of data: 100% James Preparation of manuscript 80% James 20% Thomas								
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research Candidature.								
Signed	J.Fidal		Date		October 3 2018				

Table 5.9: Statement of authorship paper:
A comparison between lumped and semi-distributed models using sub-daily data on the Gyeongancheon and Rodbourne catchments.

Chapter 6

A comparison between lumped and semi-distributed models using sub-daily data on the Gyeongan-Cheon and Rodbourne catchments.

6.1 Abstract

Numerous different types of urban rainfall-runoff models exist, each designed for different hydrological purposes. Current conclusions on which type of model to use are conflicted in the literature. The issue that is explored in this paper will be whether a simple lumped rainfall-runoff model can outperform a semi-distributed rainfall-runoff model. Two models were chosen for this study, the lumped rainfall-runoff model URMOD and semi-distributed Catchment Assessment Tool model. Two different approaches is used: the first is can a continuous data set and the second event based comparison. The results show that the lumped urban model is capable of matching and out-performing of the semi-distributed model.

6.2 Introduction

Different types of rainfall-runoff models exist, each designed for different hydrological purposes such as expanding stream flow records or predicting flows in ungauged catchments (Salvadore et al., 2015). One classification for hydrological models is how they consider spatially varying hydrological processes. The three most common model types are: (i) lumped, (ii) semi-distributed and (iii) fully distributed (Beven, 2011). Lumped models assume the entire catchment as a singular unit and do not take spatial variation into account. The advantage of these models is that they usually require less data in order to simulate hydrological processes than the other models. Semi-distributed models split a catchment into multiple lumped sub-catchments, and therefore require more detailed inputs than lumped models especially concerning catchments and river network structures, however spatial variation is taken into account. Fully distributed models split a catchment into a grid consisting of smaller cells in order to model the processes of each cell. In order to fully model the path of runoff through a catchment. The choice of using lumped, semi-distributed or distributed models is prevalent thought out the hydrological literature with arguments for and against choosing any particular model, Table 6.1 is a summery of six studies which compare lumped and distributed models.

Conclusions from the scientific literature indicate that if spatially varying observed data is available then distributed models are preferable to lumped models as shown in Table 6.1. However results also show that when catchment average data is not available, the semi-distributed models or even lumped models can perform better than distributed models Krysanova et al. (1999). One major problem with the majority of the literature stated in Table 6.1 is that all of these studies were conducted using daily data and for predominately rural catchments. The comparison of lumped and spatially distributed models in urban catchments is less clear due to a paucity of literature on the topic. Salvadore et al. (2015) concluded that urban catchments have fast dynamics such that daily data may not be able to represent the hydrological processes, and for models to capture urban catchments the application of data with a high temporal resolution is often needed. They also review a large number of existing models and find that the majority of urban lumped models are driven by daily data, which is contradictory

Models	Reference	Catchment	Result
NAM (lumped) WATBAL (Semi-distributed) MIKE SHE (Distributed)	Refsgaard and Knudsen (1996)	Three catchments in Zimbabwe using daily data (254 km^2, 1040 km^2, 1090 km^2)	Similar performance for the models. Lumped model performance is variable depending on amount of calibration data. No calibration data distributed models perform better.
SWAT (Semi-distributed) MIKE SHE (Distributed)	El-Nasr et al. (2005)	Jeker River basin in Belguim using daily data (465 km^2)	Fully distributed model performed best.
Three Hydrologiska Burans Vattenbalansavdelning (HBV) model (one lumped and two Distributed)	Krysanova et al. (1999)	Urban sub-basin of Elbe catchment using daily data (80 657 km^2)	All model structures performed good. Distributed models performed better when detailed data is applied.
Four HBV model (one lumped, two semi-distributed and one distributed)	Das et al. (2008)	Upper Neckar catchment in southwest Germany. (13 sub catchments ranging from 119.89km^2 to 454.65km^2.)	Both semi-distributed models outperformed the others.
Five HBV model (lumped, semi-distributed, grid-model without routing, grid-model with hillslope routing and a grid-model with both hillslope and channel routing.)	Li et al. (2015)	Glomma River catchment in Norway (18,932 km^2).	Increasing model complexity did improve model performance. But no differences between the full grid-models was found.
Three Sacramento Soil Moisture Accounting (SAC-SMA) (one lumped, two semi-distributed)	Ajami et al. (2004)	Illinois River basin at Watts catchment (1645 km^2)	At the catchment outlet the model performances were similar.

Table 6.1: Comparison of distributed and lumped models present in the literature

of the argument that sub-daily data must be applied to capture the effects of the urban catchments. The existing literature is limited with regards to comparisons between models at sub-daily time step. Atkinson et al. (2003) compared 8 versions of a similar model structure using both daily and hourly data, on a 47 km^2 catchment in New Zealand. The comparison used both performance criteria and stream flow hydrographs. They concluded that for hourly data the most advanced fully distributed version of the model performed the best. However, it was noted that some of the less detailed models did produce results of similar accuracy of the most detailed model. Conversely Dotto et al. (2011) compare two models: (i) a semi-distributed model (MUSIC) and (ii) a lumped model (KAREN), in order to explore parameter sensitivity. The study was conducted on five small urban catchments ranging from 0.105 km^2 to 1.056 km^2 in Melbourne, Australia with sub-daily data using a combination of performance criteria and hydrographs to obtain model performance. Conclusions indicated that both models performed equally well and were able to reproduce observed data. Hence the conclusion drawn by Salvadore et al. (2015) that distributed models are the future of urban modelling appears to be not uniformly supported by published studies. Further research needs to be conducted for a comparison between lumped and distributed urban models on sub-daily data in order to determine if and when distributed/ semi-distributed models are preferred to lumped models.

This paper expands upon current literature through a comparison of a lumped (URMOD) and a semi-distributed (CAT) model using daily and sub-daily data for two catchments. The first catchment is the the Gyeongan-Cheon river catchment located in South Korea, and the comparison explores both models performance using continuous daily and sub-daily data. The second catchment is Rodbourne located within the river Ray catchment, and draining the town of Swindon in the UK. This comparison explores the models' performance on a smaller catchment (Rodbourne) which is 5.5 km^2 and relying on an event-based comparison. The chosen models are: (i) the semi-distributed Catchment Assessment Tool (CAT) (Kim et al., 2012), developed specifically to investigate the impact of urbanisation on storm runoff and (ii) the lumped model URMOD, which is a lumped-conceptual catchment parameter-parsimonious rainfall-runoff model for use on both urban and rural catchments.

6.3 Model Description

The two models selected for this study are the semi-distributed CAT model and the lumped model URMOD. Model descriptions for both are detailed in this section.

6.3.1 Semi-distributed model: the CAT model

The CAT model is a physically-based semi-distributed water cycle analysis model developed at the Korea Institute of Civil Engineering and Building Technology (KICT) (Kim et al., 2012). Splitting a catchment into a number of smaller sub-catchments, a link-node structure is created which enables routing of runoff from sub-catchments into an entire catchment response. Hence, runoff values can be obtained from the links between sub-catchments as well as flow at a catchment outlet. The CAT model can still operate as a singular lumped model if the catchment is not split into smaller sub-catchments. CAT has a total of 31 model parameters in need of estimation or calibration, as shown in Table 6.7 in Appendix A, alongside a number of routing parameters dependent on the routing method chosen.

The CAT model requires access to hydrological, topographical, land-use and soil data to estimate the model parameters. The observed hydrological data needed are evaporation and precipitation data, which can be averaged for the entire catchment or for each sub-catchment. The model has the ability to generate estimated potential evaporation data using the Penman-Moneith method (Monteith, 1965), this is instead of inputting already calculated potential evaporation.

Next, spatial catchment properties are needed such as topographical, land-use and soil data. The topographical data is represented by a digital elevation model (DEM) from which slope data can be derived. The land-use data needed is impervious and pervious percentage land cover and the percentage of vegetation (e.g. crops, gardens) within the pervious areas. The last set of data is soil property data, this consists of soil depth, saturated soil moisture, residual soil moisture, wilting point, soil moisture content at field capacity, soil moisture content at wilting point, saturated hydraulic conductivity and saturated horizontal hydraulic conductivity. The CAT model's parameters are physically-based parameters and should ideally be specified appropriately based on the physical catchment data. When such data is not available and estimation is not possible, observed stream flow data is required in order to calibrate the models parameters. Calibration of these parameters can be undertaken though a trial-and-error approach in order to obtain simulated stream flow data which can be compared to observed data, or by applying an objective function method to do the same process but automated. Figure 6-1 shows all of the CAT models main processes alongside any sub-processes.

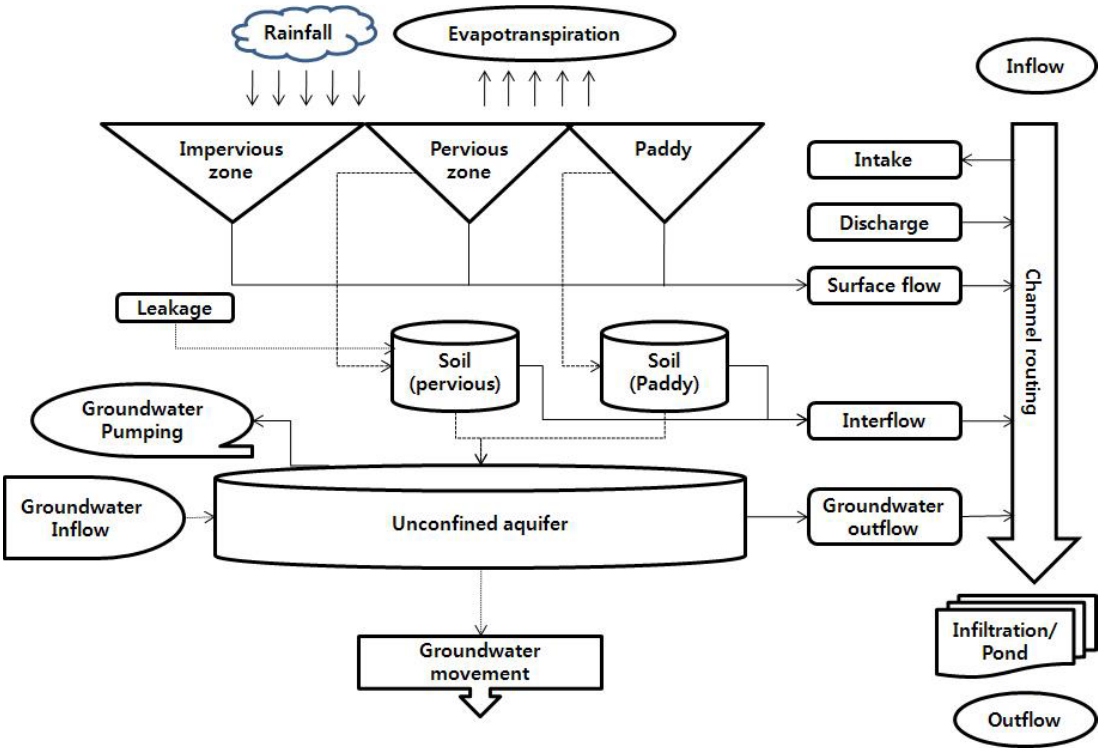


Figure 6-1: Full model diagram of the water cycle processes of the CAT model (taken from Kim et al. (2012) and the Catchment Hydrologic Cycle Assessment tool User Guide)

The CAT model's main processes are: (i) infiltration, (ii) groundwater flow, (iii) evaporation and (iv) channel routing. CAT has access to three infiltration methods: (i) the rainfall excess method based on vertical/horizontal hydraulic conductivity (Chow and Mays, 2010), (ii) the Green and Ampt method (Green and Ampt, 1911), and (iii) the Horton method (Horton, 1933). Groundwater flow between sub-catchments is controlled via the Darcy equation. Four different methods can be applied for channel routing: (i) the Muskingum (Linsley et al., 1949), (ii) Muskingum-Cunge (Cunge, 1969), (iii) Kinematic wave method (Lighthill and Whitham, 1955), or (iv) no-routing method, such that excess rainfall is converted directly into runoff. A different amount of the main processes will be used later on in the chapter.

6.3.2 A lumped model: URMOD

URMOD is a lumped conceptual catchment parameter-parsimonious rainfall-runoff model with nine parameters in need of calibration. All of URMOD's parameters are shown in Table 6.8 in Appendix B, and if required observed hydrological data in order to calibrated the parameters. The hydrological data consist of observed rainfall, runoff and potential evaporation data. URMOD also needs the total area of the catchment and percentage of the catchment which is urbanised, such as the URBEXT value (Bayliss et al., 2006). To calibrate the nine free parameters URMOD uses the shuffled evolution complex algorithm (Duan et al., 1993). Once calibrated the set of parameters obtained are applied to the model alongside observed rainfall and evaporation data in order to generate simulated runoff data at a catchment outlet. Figure 6-2 shows all of URMODs main processes.

The two main processes of URMOD are: (i) infiltration and runoff generation, and (ii) channel and pipe routing. URMOD generates two contributions for runoff, one from the urban areas and one from the rural areas. The model's infiltration and runoff generation is based on a conceptual soil column approach, such that the precipitation that does not infiltrate is turned into runoff. The soil moisture level is controlled by three processes, (i) infiltration into the soil column, (ii) drainage and (iii) evaporation.

Infiltration in the rural areas is only dependent on the moisture content in the soil column, whereas the infiltration from the urban areas is dependent on both the soil moisture content as well as a calibrated term γ, accounting for the restricted infiltration on urban surfaces. If $\gamma = 1$ then no infiltration occurs on the urban surfaces, and if $\gamma = 0$ then infiltration is equal to the rural infiltration. Precipitation that does not infiltrate into the soil column is turned into direct runoff. Two direct runoff values are obtained, one from the rural area and one from the urban area. The model's second process is the routing of the direct

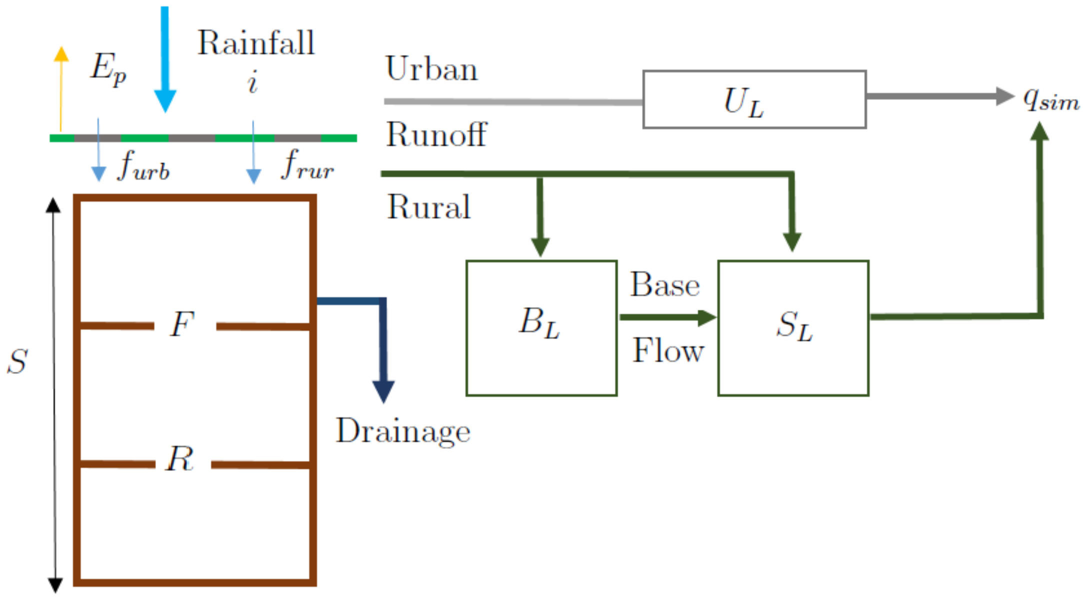


Figure 6-2: Full model diagram of URMOD, with conceptual soil column and zones on the left hand side. Urban pipe routing system and rural routing system on the right hand side.

runoff, both are routed separately in two parallel linear reservoirs and then rejoin at the catchment outlet for the total flow. Routing of runoff from the rural areas is achieved using a linear reservoir model with a proportion of the runoff first routed through a local base flow reservoir before re-emerging and then routed through a surface flow reservoir. The proportion of runoff designated as surface flow bypasses the baseflow reservoir and is routed directly through the surface flow reservoir. The baseflow and surface flow is then combined to become the rural runoff at the catchment outlet. The urban runoff is routed through a bounded pipe with an overflow mechanic, such that if the capacity of the bounded pipe is reached, then the excess runoff is routed through the rural routing. Finally the urban runoff is then combined with the rural runoff to become the total runoff at the catchment outlet.

6.4 Case study: The Gyeongan-Cheon river Catchment and the River Ray sub-catchment Rodbourne

The 260.1 km^2 Gyeongan-Cheon River Basin is located in the northwestern area of South Korea and is split into 11 sub-catchments as shown in Figure 6-3. The 11 sub-catchments were chosen based on river outlets and the transportation of flow through the subcatchments. Urban land-cover accounts for approximately 19% of the total catchment area. The catchment is equipped with a singular flow monitoring station and rainfall gauge.

Figure 6-3: Map of Korea and the Gyeongan-Cheon River Basin with subcatchments.

The second catchment is Rodbourne, located within the river Ray catchment (National River Flow Archive (NRFA) number 39087) which contains the town of Swindon as shown in Figure 6-4.

The river Ray catchment is approximately 84.1 km^2 and contains Swindon, a town located in Wiltshire 78 miles west of London, with a population of approximately 182,000. The 5.5 km^2 Rodbourne catchment located within the river Ray catchment and is highly urbanised with approximately 46% of the area impervious. The catchment has a singular rainfall gauge and flow monitoring station.

6.4.1 Catchment data

Gyeongan-Cheon Catchment

Hydrological data

The data for this study was provided by the Korea Institute of Civil Engineering and Building Technology. Two sets of hydrological data were used. The first set consists of average observed (i) daily precipitation (i), (ii) river flow (q_{obs}), and (iii) potential evaporation (E_p) spaning 10 years from 01-01-1998 to 31-12-2007. The second set consists of (i) average observed hourly precipitation, (ii) river flow and (iii) climate data (including: average temperature, wind, humidity and solar data) spanning 11 years from 01-01-1998 to 31-12-2008. This data was then converted into three hourly and twelve hourly time steps for the purpose of the study. The precipitation data was originally obtained from Suwon, Icheon and Yangpyeong climate stations. The evaporation data were originally obtained in the form of climate data (average temperature, minimum and maximum temperature, wind speed, humidity and solar radiation) and potential evaporation calculation using the Penman-Monteith method. The river flow data was originally collected by the Han River Flood Control Centre. No missing observations were present in the daily data.

Figure 6-4: Map of River Ray catchment with Rodbourne highlighted in black

The rainfall data showed strong seasonality. Average precipitation during the dry winter months October-March was below 35 mm accounting for 9.9% of yearly rainfall. As a result the flow data reflects this with certain months had very low or missing flow data during the dry winter months. In contrast, during the humid summer period April-September average monthly precipitation was above 75 mm with the total rainfall in the wet months accounting for 90.1% of annual rainfall, with July and August having total precipitation values as high as 410 mm and 329 mm, respectively. The results section will focus on July and August as every year of the observed data for these two months had no missing rainfall data but potentially some missing flow data and will test the models abilities to simulate extreme rainfall. During calibration of URMOD the dry season will be omitted. Water balance checks were made for both the Gyeongan-Cheon and Rodbourne catchments to make sure neither catchment had large discrepancies. Both water balance checks were deemed to be acceptable with both catchments having extra water which is too be expected due to human influence on the catchment.

Land use data

Land-use data was obtained to estimate CAT models parameters. A digital elevation model (DEM) was obtained from the National Geographic Information System (NGIS) digital map, with scale 1:5000, and used to determine slope values for each sub-basin. Also derived by DEM, catchment boundary and average impervious area ratio was obtained via land use distribution map from the Korean Ministry of Environment with scale 1:25000. Vegetation area ratios were extracted from a previous study Jia (1997). Soil data was obtained based upon soil maps from the Agriculture Science technology information system of Korea in 1:25000 scale, then the soil texture values were obtained from Rawls and Brakensiek (1985). The hydraulic conductivity of the riverbed material and channel properties were obtained from the Gyeongan-Cheon Fundamental Planning Report for River Improvement.

Rodbourne Catchment

The hydrological data obtained for this study was used in Miller et al. (2014) and provided by one of the authors (Kim, Hyeonjun, (pers.comm.) 2017). Details of how the data were obtained is outlined here, but for full details see the aforementioned paper. The data include 15 minute observed: precipitation (i), river flow data (q_{obs}) and potential evaporation (E_p), spanning from 01/05/2011 to 17/10/2012 in 15 minute intervals. All of the data was converted to hourly data for the purposes of the study. Precipitation data was measured using a singular tipping bucket located within the Rodbourne catchment. Flow monitoring data is recorded at 1 minute resolution but was averaged to 15 minute and then to hourly data for the purposes of this study. Evaporation data is a contribution of two data sets, the first set spanned May 2011 to April 2012 and was measured from a eddy covariance mast at a monthly period locally in a peri-urban area of housing Ward et al. (2013). Additional data was obtained from the Met Office Rainfall and Evaporation Calculating System (MORECS) spanning May 2011 to October 2012. The local measured data revealed that the evaporation was 60% of the MORCES value. The remaining May 2012 to October 2012 potential evaporation data was obtained by scaling the 60% up the MORECS data.

A total of 15 individual events were selected, based on peak flow and total rainfall, such that the total rainfall had to exceed 10 mm, and peak flow over 0.5 m^3s^{-1}. One event was under 0.5 m^3s^{-1} but the total rainfall exceed 10 mm so it was included (Event 2). The length of the rainfall events were also taken into account to select events which were both short (i.e less than a day) as well as extended periods of more than a day of rain. Table 6.2 shows a summary of the events, including: (i) the start date time, (ii) duration, (iii) the observed peak flow and (iv) total rainfall.

In order to determine the percentage of urban land-cover in a catchment for CAT and URMOD, the URBEXT_{2000} catchment descriptor Bayliss et al. (2006) was used. The subscript 2000 denotes that the $50m \times 50m$ land-cover data used to construct the index refers to the period between the years of 1998-2000. URBEXT_{2000} uses a contribution of both urban and sub-urban land-cover classes, with the urban land-cover consisting of roofing, roads and man-made structures, whereas the sub-urban section is a mix of vegetation and semi-built up areas. Only half of the sub-urban area is counted as urban as it is assumed that half of the sub-urban is made up of vegetation e.g (gardens or parks) (Bayliss et al., 2006).

Event Number	Start date	Time	Duration (hr)	Total Rainfall (mm)	Peak flow (m^3s^{-1})
1	2011-06-12	05:00	19	17.4	0.892
2	2011-12-13	23:00	39	10.2	0.637
3	2012-01-03	01:00	22	15.8	0.989
4	2012-05-01	02:00	21	15.2	1.426
5	2012-06-23	18:00	21	17.4	0.707
6	2012-07-12	15:00	20	16.2	1.083
7	2012-08-15	10:00	15	16.4	1.133
8	2012-08-29	09:00	23	21.2	0.969
9	2012-09-23	10:00	37	29.2	0.844
10	2012-10-04	18:00	44	28.4	0.741
11	2012-04-27	09:00	68	33.4	1.064
12	2011-08-07	11:00	17	14.6	0.462
13	2012-06-30	06:00	32	14	0.716
14	2012-08-02	11:00	65	22.2	0.916
15	2011-12-11	15:00	42	25	0.971

Table 6.2: 15 events selected for Rodbourne analysis

6.4.2 Experimental design

CAT calibration method

As indicated in Section 6.3.1 the CAT model has 31 physically-based catchment parameters that should be estimated from the catchment data, alongside a total of 9 routing parameters. The actual number of the parameters used for routing is dependent on the routing method applied. The Muskingum has three, Muskingum-cunge have four parameters and Kinematic has six parameters. Two sets of parameters were obtained for the Gyeongan-Cheon catchment, one set for the daily data and the second set for the sub-daily data. All except eight of the catchment parameters were obtained through the catchment data. The remaining eight parameters were obtained from a previous study on the Gyeongan-Cheon catchment (Jang et al., 2016) and estimated using an objective function on the years (1999-2001). The same non-routing parameter set was applied to the sub-daily analysis in this study due to the parameters not temporal related.

The routing method used in this study was the Muskingum method, which resulted in three new parameters in need of estimation: (i) ΔT the time step, (ii) K storage coefficient of the ratio of the storage to the outflow, and (iii) a dimensionless constant x for the relative importance of the inflows and the outflow. The daily routing parameters were adopted from previous studies. The 12-hourly routing parameters were estimated based on a trial-and-error approach, relative to the daily data. Due to the size of the catchment and that flow takes less than 3 hours to travel through the catchment an alternative routing strategy was applied for the 3-hourly and hourly time step. Such that an instantaneous routing method was used in order to show that the flow does not need to be delayed.

Similar to the Gyeongan-Cheon catchment, the parameters for Rodbourne catchment were obtained from a previous study (Miller et al., 2014). During the previously mentioned study all parameters except for impervious cover had to be estimated, through trial-and-error calibration using data from 5 of the events in Table 6.2 (Event 1,4,5,7,8). The same parameters were used for this thesis except for the routing parameters due to a different time step being used. In order to calculate the routing parameters a trial-and-error approach was used, using 3 events from Table 6.2 (Event 1,4 and 5), two short events and one extended event.

URMOD calibration method

Unlike CAT, all of URMOD's nine free parameters require calibration. A calibration period needs to be defined, as outlined in Klemeš (1986), and a split sample method is suggested such that the calibration and validation do not overlap. The calibration and validation method used by URMOD is an advancement to the original split-sample method defined as a systemic jackknife re-sampling method. Since the calibration scheme is using individual events initial conditions need to be considered. As initial conditions will vary depending on the time of year or soil moisture after rainfall. This was considered for this thesis but only the normal initial conditions was used, this is to test the models stop/ start performance instead of continuous. The purpose of using two different calibration strategies (yearly and event-based) was too test how URMOD performs under experiments it was not initially designed for. A description of how the jackknife method will be applied to both Gyeongan-Cheon and Rodbourne catchments will be presented.

A total of 15 rainfall events are selected as shown in Table 6.2. Event 1 is used to calibrate URMODs nine parameters in order to obtain a parameter set θ_1. Once calibrated θ_1 will be applied to URMOD in order to simulate runoff for each of the remaining 14 events. A performance criterion can be obtained by comparing the simulated and observed data for each of events resulting in 14 performance criteria $Z_{1,k}$ $k = 1, ..., 14$ (one for each event except event 1). Once completed event 2 is used to calibrate URMOD in order to obtain a different parameter set θ_2. 14 new performance criteria can be obtained for the remaining 14 events $Z_{2,k}$ $k = 1, ..., 14$. This process is then repeated systematically, each time using a different event until 15 parameter sets are obtained $\theta_j, j = 1, ..., 15$ and 14 performance criteria for each event $Z_{j,k}, k = 1, ..., 14, j = 1, ..., 15$ is obtained. A median of each $Z_{j,k}$ for each catchment can be obtained resulting in a single performance criterion for each event. $\tilde{Z}_j$.

A similar method is applied to the data from the Gyeongan-Cheon catchment but rather than calibrating on individual events, the calibration period is the wet months defined in 6.4.1. For the daily data a two-year jackknife calibration method is used such that the ten years of data are split into five two-year periods. Whilst the sub-daily data uses a yearly jackknife method such that the eleven years of data will be split into eleven yearly periods.

Firstly the data set is split into five equal length non-overlapping two-year periods starting at the year 1998. The years 1998 and 1999 are removed from the data set, to calibrate URMOD resulting in a parameter set θ_1. Once calibrated θ_1 will be applied to URMOD in order to simulate the remaining eight years of data. A performance criteria will be obtained by comparing the simulated and observed data for the remaining data for every year, resulting in eight performance criteria $Z_{1,k}, k = 1, ..., 4$. This process is repeated systematically, each time using a different set of two years, until five parameter sets $\theta_j, j = 1, ...5$ and four performance criteria for each of the sets $Z_{j,k}, j = 1, ..., 5, k = 1, ..., 8$ are obtained. Similar to Rodbourne an median performance criteria will be obtained for each set of two years resulting in five performance criteria. The sub-daily data will follow a similar process, but will use a yearly jackknife method and obtaining performance criteria for each July and August. This results in 11 parameter sets $\theta_j, j = 1, ...11$ and 22 performance criteria $Z_{j,k}, j = 1, ..., 11, k = 1, ..., 10$.

Model performance criteria

Two performance criteria will be used in this study, the Nash-Sutcliffe efficiency statistic (NSE) and the Mean Absolute Error (MAE). The two performance criteria are chosen as they measure different aspects of the model performance. The NSE is sensitive to peaks due to the squares of the difference between observed and simulated runoff, whereas the MAE takes the absolute value of the difference. The Nash-Sutcliffe efficiency (NSE) statistic Nash and Sutcliffe (1970) is defined as:

$$\mathrm{NSE} = 1 - \frac{\sum_{t=1}^{n}(q_{obs} - q_{sim})^2}{\sum_{t=1}^{n}(q_{obs} - \bar{q}_{obs})^2} \tag{6.1}$$

where q_{sim} is the simulated river flow from either CAT or URMOD and $\bar{q}_{obs}$ the mean of the observed river flow. The range of NSE values falls between one and $-\infty$, with a value of one indicating perfect fit, i.e $q_{obs} = q_{sim}$. The second performance criteria is the Mean absolute error MAE defined as:

$$\mathrm{MAE} = \frac{\sum_{t=1}^{n}|q_{obs} - q_{sim}|}{n} \tag{6.2}$$

with zero indicating a perfect fit between observed and simulated values. Values of the MAE range from zero to ∞. For consistency when stating performance criteria the subscript u will be used to indicate URMOD and c will be used to indicate CAT, e.g. MAE_u and MAE_c.

Binomial hypothesis test

To determine the relative model performance, a Binomial hypothesis test approach is used. The test utilises a success/failure approach when comparing the performance criteria of the two models; either CAT outperforms URMOD ($\mathrm{NSE}_c > \mathrm{NSE}_u$), success or the reverse (failure). The case when model performance is exactly equal ($\mathrm{NSE}_c = \mathrm{NSE}_u$) is not taken into account, as this is considered an unlikely occurrence.

Define a trail as the difference in performance criteria for both models for a singular event (NSE_c - NSE_u). A hypothesis test can be set up such that the null hypothesis (H_0) assumes that for a specified number of trials there is no difference in model performance, and therefore the probability that either model outperforms the other can be assumed to be a half:

$$H_0 : p = 0.5. \tag{6.3}$$

The alternative hypothesis (H_1) can be set up as a two-tailed test, such that two test's will be performed one looking for URMOD outperforming CAT and vice versa. Hence the alternative hypothesis can be given as:

$$H_1 : p \neq 0.5. \tag{6.4}$$

Now let c denote a set of trials, define a success as CAT outperforming URMOD ($\mathrm{NSE}_c > \mathrm{NSE}_u$) whereas a failure is URMOD outperforming CAT ($\mathrm{NSE}_c < \mathrm{NSE}_u$). Let V be a random variable defined as the number of trials of CAT outperforms URMOD, i.e. the number of successes obtained in h trials. Thus the probability of $V = v$ successes is given as a binomial distribution $V \in B(c, p)$ with a cdf given as:

$$P\{V = v\} = \binom{c}{v} p^v (1 - p)^{(c-v)}. \tag{6.5}$$

The hypothesis test can be formed for a predefined significance level e.g $\alpha = 5\%$. The observed number of success v is compared to the critical interval defined as $v \leq B(c, p)_{\frac{\alpha}{2}}$ and $v > B(c, p)_{1-\frac{\alpha}{2}}$ are the $(\alpha/2)$ and $(1 - \alpha/2)$ quantiles of the binomial distribution. If v falls within the critical interval then the null hypothesis can be rejected such that there is a difference between CAT and URMOD performance in favour of one of the models. But if v does not fall within the critical interval then the null hypothesis can be accepted such that there is no difference between CAT and URMOD.

The binomial test will be applied to compare model performance for sub-daily data for the Gyeongan-Cheon catchment and compare model performance on the events for Rodbourne. For both catchments a two tailed test will be performed. For the Gyeongan-Cheon catchment each trial will be defined as a comparison of performance criteria of CAT and URMOD for both July and August, resulting in $c = 22$. Whereas for Rodbourne a trail will be defined as the difference in performance of the performance criteria for an event, resulting in $c = 15$.

6.5 Results

6.5.1 Assessing performance of URMOD and CAT on the Gyeongan-Cheon catchment

This section explores the difference in performance between URMOD and the CAT model on the Gyeongan-Cheon catchment. Firstly using daily data then using 12-hourly, 3-hourly and hourly time step data.

Comparison of URMOD and CAT using daily data

As outlined in the calibration method of URMOD in Section 6.4.2, URMOD is calibrated four times on two year periods and then performance criteria is obtained for each of the non-calibration years. This process results in three performance criteria for each year, the median is then taken for each year. Figure 6-5 shows the difference in median performance criteria for both the NSE and MAE for the daily data sub-setted by year. The left hand figure (a) is the difference in NSE, reported as NSE_c -NSE_u and the right hand figure (b) is the difference in MAE, it is reported as MAE_u - MAE_c. Positive values indicate CAT has greater performance than URMOD, shows on Figure 6-5 as the black circles. In contrast, negative values indicate that URMOD performs better than CAT, represented as red triangles.

The results in Figure 6-5 show that URMODs performance is better than the CAT model when representing the daily data when the MAE performance criteria is applied. Whilst all of the MAE values obtained by both models were large (> 2), when looking at Figure 6-6 flow exceeds 200 $\frac{m^3}{s}$ hence larger values of the MAE are within acceptable ranges. However the reverse is true when the analysis is based on the NSE criteria, with values greater than 0.2 for URMOD and 0.4 for CAT. There are periods even during the wet months when there is little rainfall, and as such the CAT model does not account for zero rainfall on the baseflow, whereas URMOD can account for the baseflow. Whilst CAT does a groundwater reservoir to account for baseflow during low rainfall, due to the parameters of the model it does not and as such creates less baseflow than observed. However for the NSE which takes a squared value this puts an emphasis on the peaks, indicating that the CAT model does perform better when

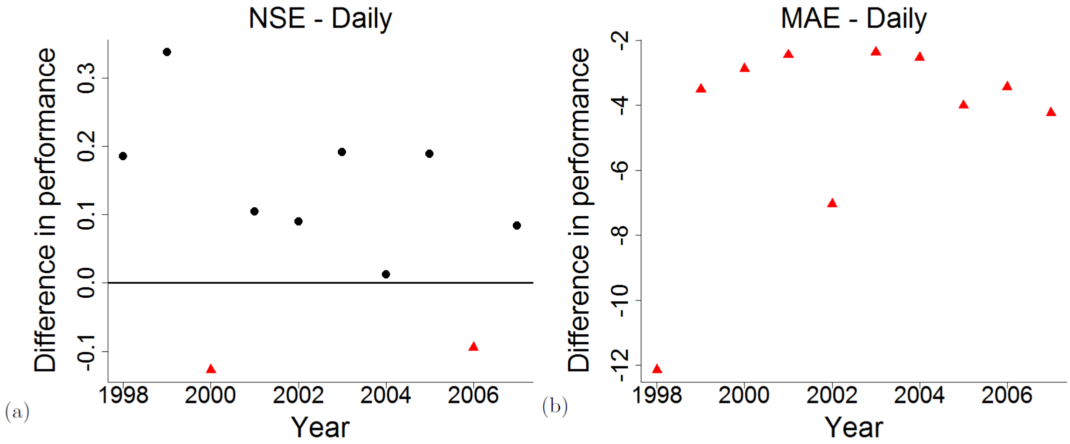


Figure 6-5: Comparison of URMOD and CAT performance criteria values using daily data on the Gyeongan-Cheon catchment. Left hand figure (a) is the difference in performance of the NSE and the right (b) is the difference in performance of MAE. Positive values show CAT has performed better than URMOD, whilst red triangles show URMOD has performed better than CAT.

exploring peak performance, this is highlighted in Figure 6-6, since URMOD under estimates peak flows more than CAT over estimates.

Figure 6-6 shows the simulated flow by both URMOD and CAT alongside the observed river flow and observed rainfall for the year 2005. The parameters used for URMOD were calibrated on the years 2006-2007. The results show that CAT is prone to over estimation whilst URMOD tends to under-estimate peak flow. Base flow is low and as such no conclusion can be drawn for either model. One observation from Figure 6-6 is that during the first large rainfall event (July) the CAT model is delaying the simulated flow until after the observed flow resulting in a longer routing period than needed. Whereas URMOD simulated flow is at a similar time to the observed flow indicating routing is accurate.

Comparison of URMOD and CAT using subdaily data

This section explores the difference in model performance between the URMOD and the CAT models on the Gyeongan-Cheon catchment when using sub-daily data. The results compares the NSE values for both the URMOD and the CAT models in the form of figures. The MAE is not be presented further as similar to the daily results as error is large across all the years. As stated in Section 6.4.2 the

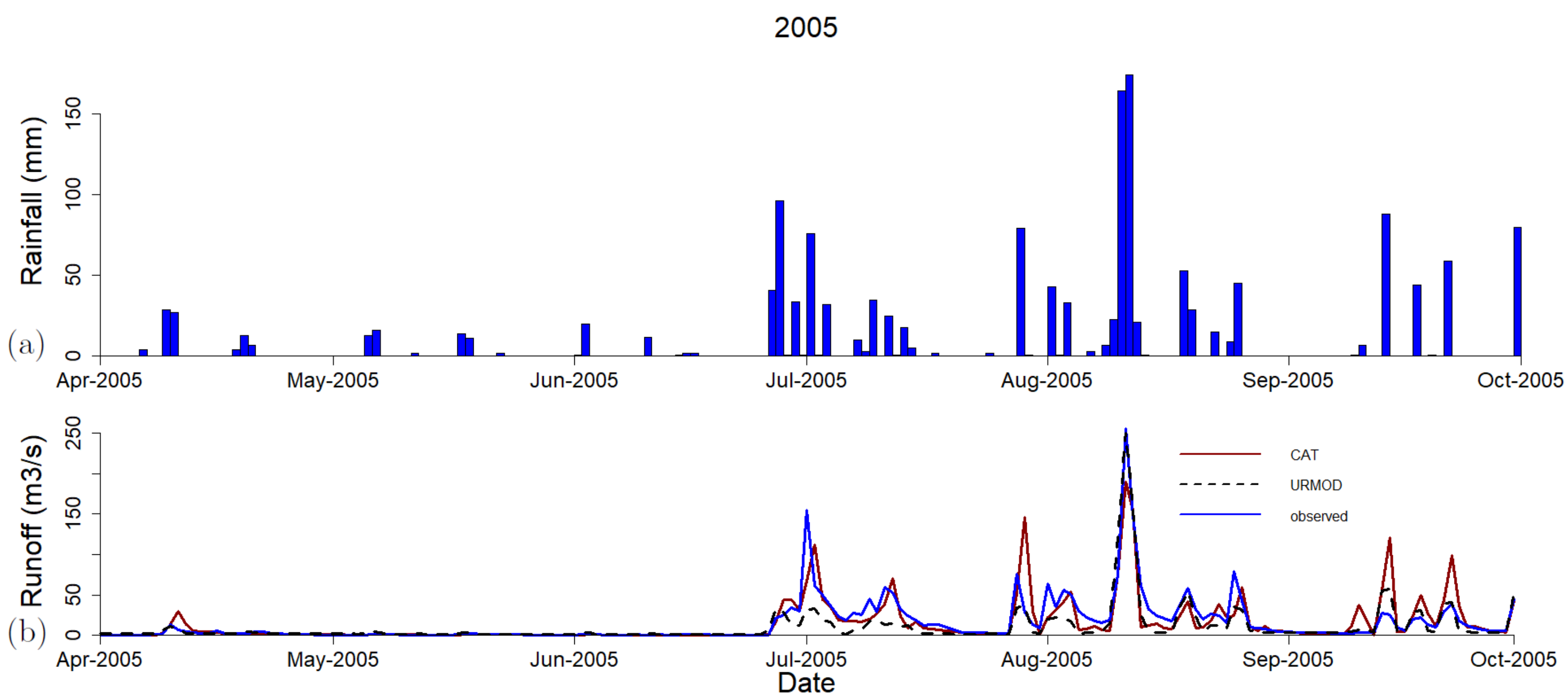
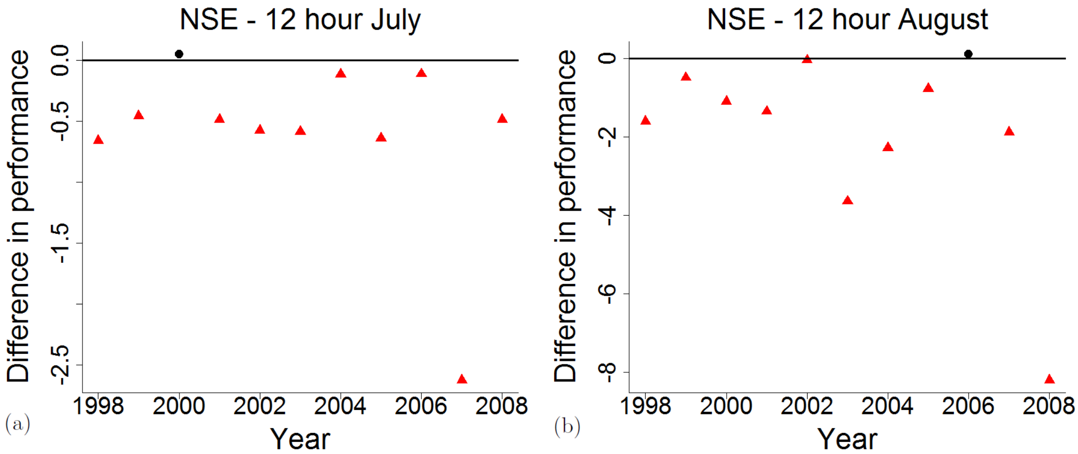
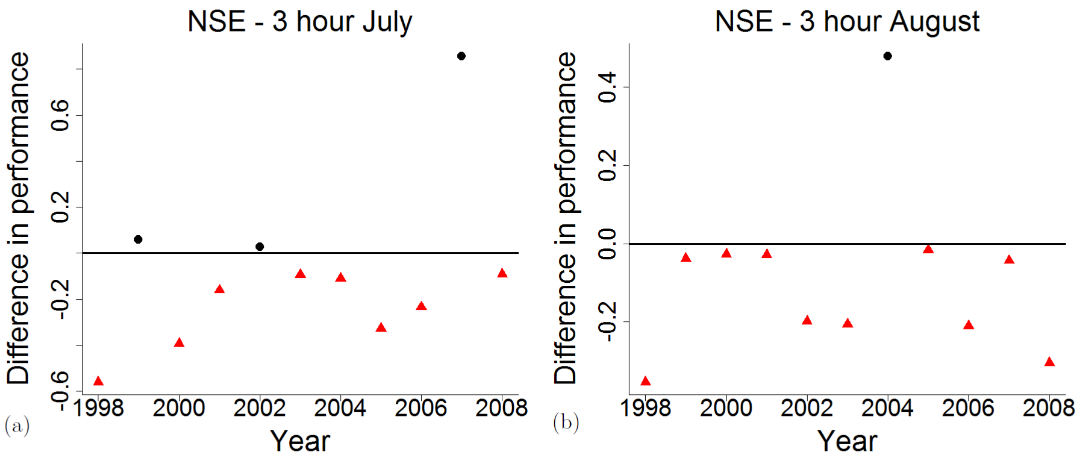


Figure 6-6: Hydrograph of simulated CAT and URMOD on the wet season 2005. Top figure (a) is observed rainfal (mm). Bottom figure (b) is observed river flow (solid blue line), estimated CAT river flow (solid red line) estimated URMOD river flow (dashed black line)

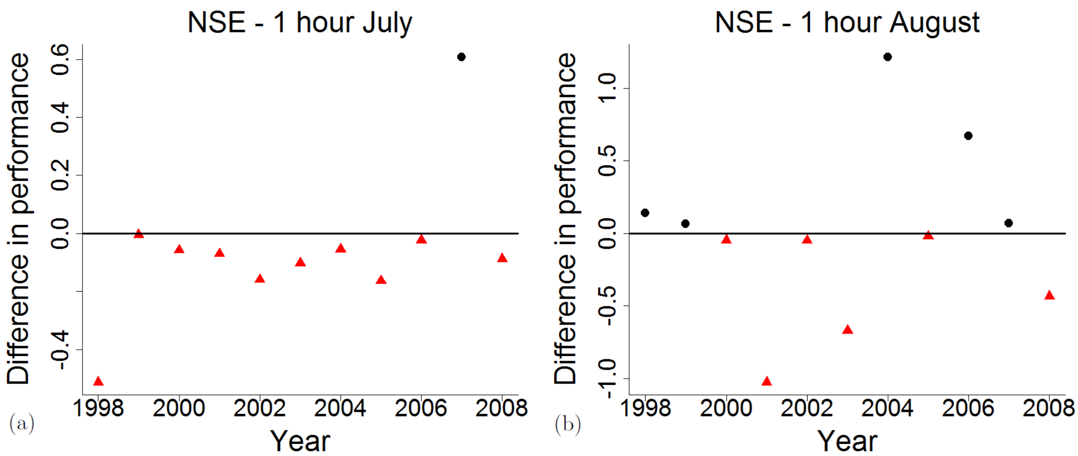
analysis focuses on the two wet months of July and August. As detailed in Section 6.4.2 due to the calibration method applied to URMOD, multiple performance criterion is obtained for a single month, the median of the performance criterion is obtained and this value is compared with the equivalent performance criteria obtained by the CAT model. This section also uses the binomial hypothesis test presented in Section 6.4.2 to assess if the CAT or the URMOD model consistently perform better than the other. The number of trials will be $h = 22$ (11 years of both August and July) and the predefined significance level will be $\alpha = 5\%$. The binomial hypothesis test will be set up as a two tailed test.



(a) Comparison of URMOD and CAT NSE values using 12-hourly time step data.



(b) Comparison of URMOD and CAT NSE values using 3-hourly time step data.



(c) Comparison of URMOD and CAT NSE values using hourly time step data.

Figure 6-7: Comparison of URMOD and CAT NSE values. The black dots show that CAT has performed better than URMOD, whilst the red triangles show URMOD has a larger NSE than CAT.

Figure 6-7 shows the difference in performance criteria between URMOD and the CAT model. The top figure (a) is the difference in performance of URMOD and the CAT model using 12-hourly data, the middle figure (b) is using 3-hourly data and the bottom figure (c) is using hourly data. The left hand figures (a) shows the difference in performance criteria for July whereas the right hand figures (b) shows the results for August. The black circles indicate that a larger NSE value is obtained using the CAT model than URMOD ($\text{NSE}_c > \text{NSE}_u$), whereas the red triangles show that URMOD has a larger NSE than CAT ($\text{NSE}_u > \text{NSE}_c$). Figure 6-7 shows that as the time step gets smaller, the number of occasions when URMOD has a larger NSE than the CAT model decreases, going from 20 (12-hourly), 18 (3-hourly) and 16 (hourly). Table 6.3 shows the average NSE of both URMOD and the CAT model for both July and August. One value for 3-hourly August was removed as it was skewing the average.

Time step	Average URMOD NSE (July)	Average URMOD NSE (August)	Average CAT NSE (July)	Average CAT NSE (August)
12-hourly	0.61	0.34	0.003	-1.59
3-hourly	0.58	0.36	0.45	0.3
hourly	0.5	0.15	0.44	0.18

Table 6.3: Average performance criteria for July and August NSE for both URMOD and CAT.

The results in Table 6.3 show that the average NSE for URMOD decreases as the time step decreases. However for the CAT model the average NSE increases from the 12-hourly to the 3-hourly time step for both July and August, but for the hourly time step the average NSE for July is similar but the August average NSE decreases. The increase in performance between 12-hourly and 3 and 1-hourly results could be a result of the routing method used. Since the 12-hourly data was still using the Muskingum approach, whereas for the 3 and 1-hourly data an instantaneous value was taken.

Figure 6-8 shows the simulated flow by both URMOD and CAT alongside the observed river flow and observed rainfall for the two months July and August for the 2005 when 12-hourly data is used, Figure 6-9 is using 3-hourly and Figure 6-10 is using hourly data. The parameters for URMOD were calibrated on the

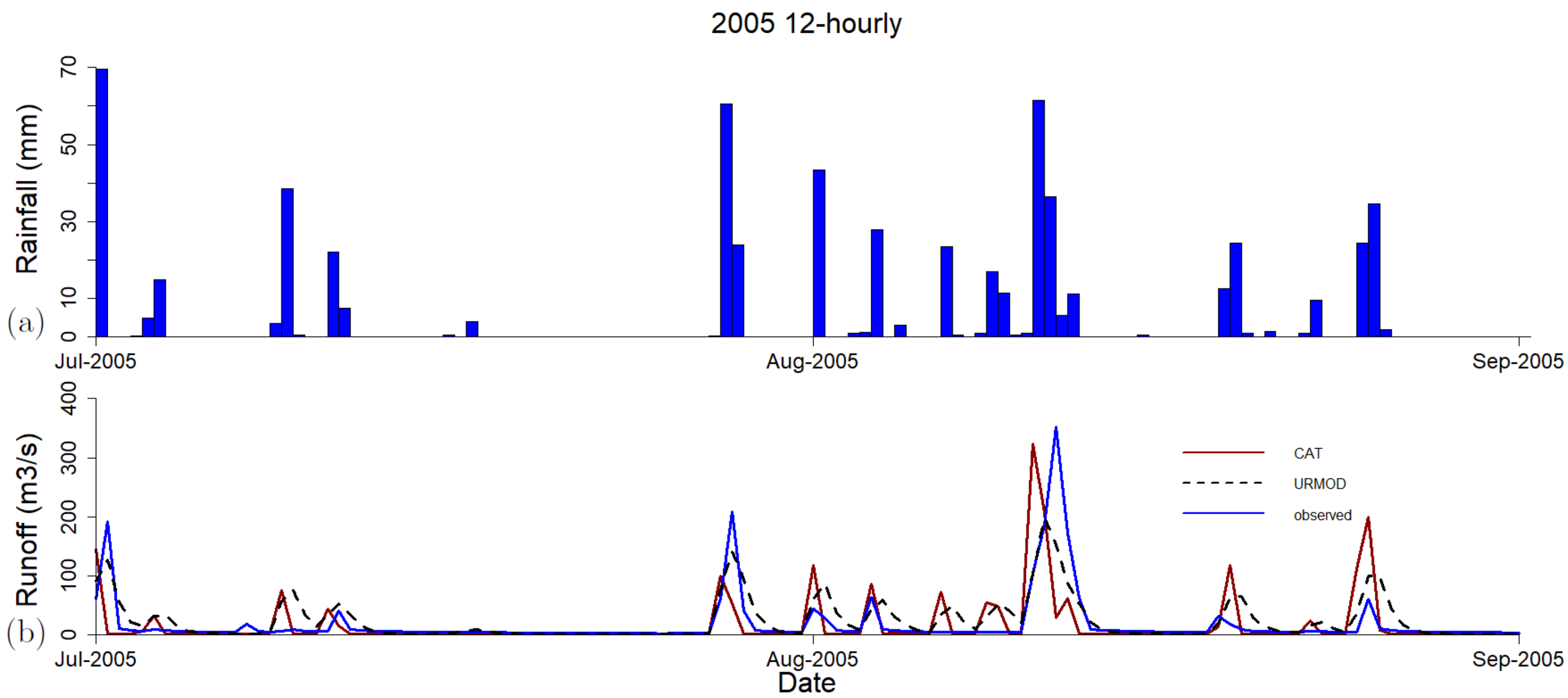


Figure 6-8: Hydrograph of simulated CAT and URMOD on the months July and August in 2005 using 12-hourly data. Top figure (a) is observed rainfal (mm). Bottom figure (b) is observed river flow (solid blue line), estimated CAT river flow (solid red line) estimated URMOD river flow (dashed black line)

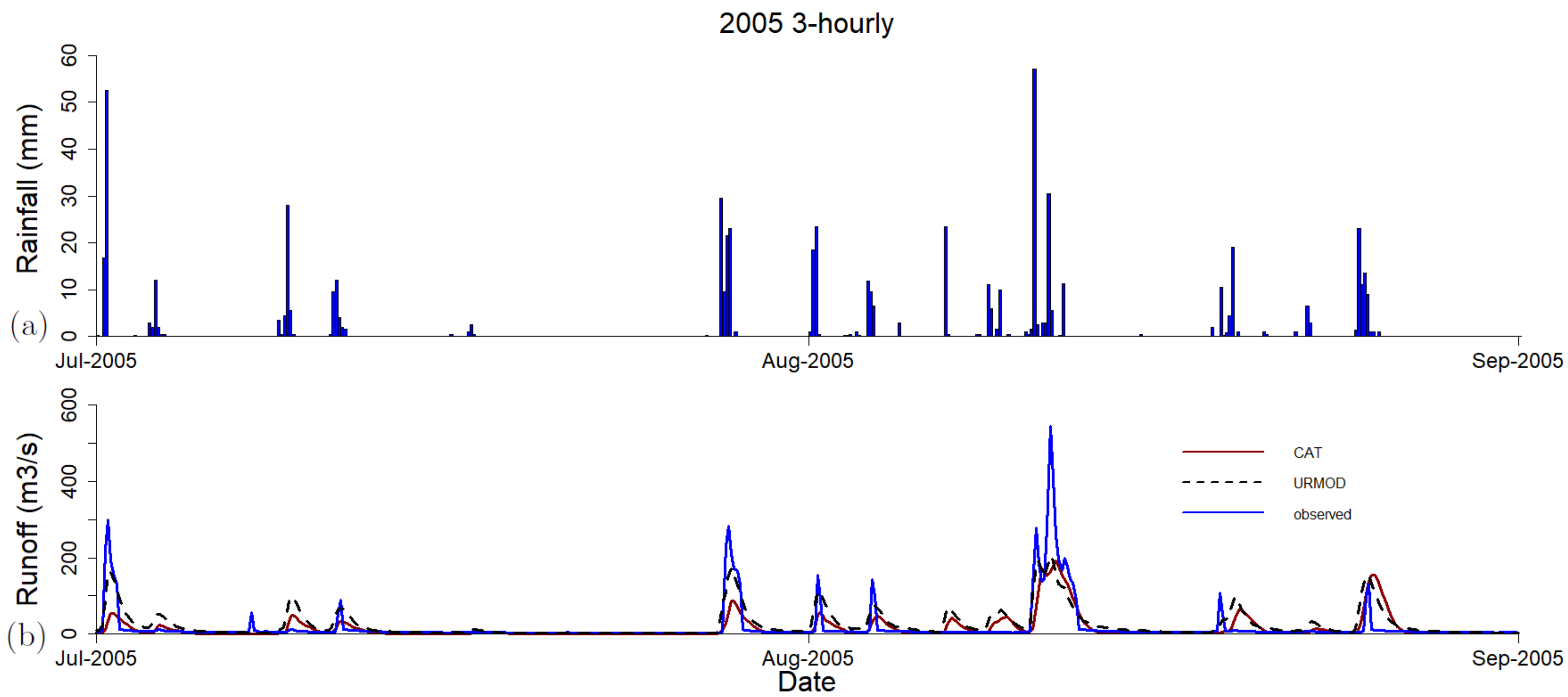


Figure 6-9: Hydrograph of simulated CAT and URMOD on the months July and August in 2005 using 3-hourly data. Top figure (a) is observed rainfal (mm). Bottom figure (b) is observed river flow (solid blue line), estimated CAT river flow (solid red line) estimated URMOD river flow (dashed black line)

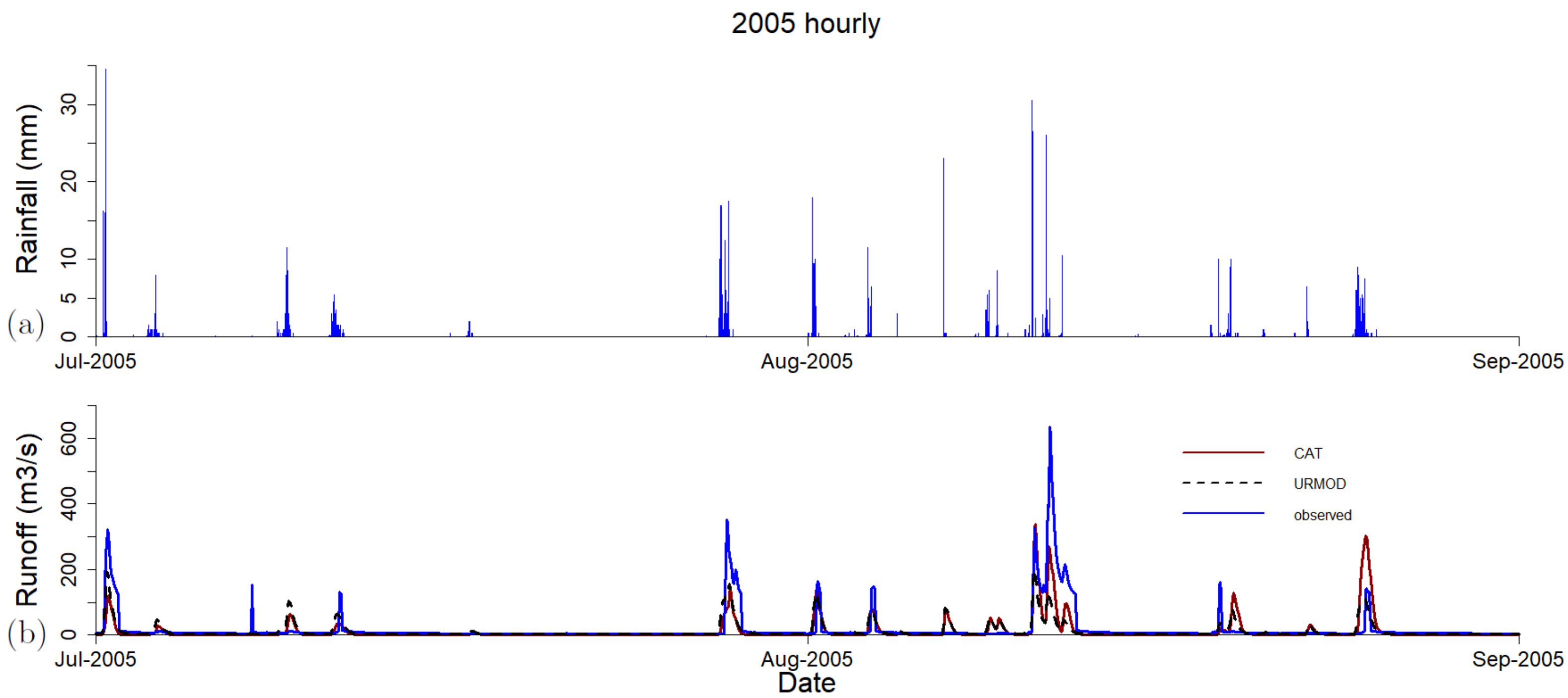


Figure 6-10: Hydrograph of simulated CAT and URMOD on the months July and August in 2005 using hourly data. Top figure (a) is observed rainfal (mm). Bottom figure (b) is observed river flow (solid blue line), estimated CAT river flow (solid red line) estimated URMOD river flow (dashed black line)

year 1999 for the three figures presented. As shown across the three figures as the time step gets smaller, the performance of the models decreases as expected. This is due to the level of detail of the flow data increasing from more data points. Figure 6-8 shows that both models do over and under simulate the observed flow, however URMOD does not do so as much as the CAT model, except for the event in mid August. As the time step decreases to 3-hourly data URMOD does still under simulate slightly but the CAT model under simulates more and both models are incapable of modelling the event in mid August. One reason for this is during the calibration due to the amount of data points the algorithm prioritises the low flow instead of the peak event. Finally for hourly data neither model performs well, with both models under simulating and at times have a lag time which is too large.

Binomial hypothesis test results

Table 6.4 shows the results of the binomial hypothesis test for each time step, the H_0 assumes equal performance whereas the H_1 assumes unequal performance in favour of either URMOD or the CAT model.

Time step	p-value	outcome
12	$5.5\text{x}10^{(-5)}$	H_0 is rejected such that there is a significant difference in model performance in favour of URMOD.
3	0.004	H_0 is rejected such that there is a significant difference in model performance in favour of URMOD.
1	0.054	H_0 can be accepted such that no significant difference in model performance can be detached.

Table 6.4: Gyeongan-Cheon catchment binomial hypothesis test.

For both 12-hourly and 3-hourly data the H_0 is rejected in favour of the URMOD, however for hourly data the H_0 is accepted such that there is no significant difference between model performance. Whilst the CAT model does not significantly perform better on any time step, the NSE results show in table 6.3 that URMOD performance does decrease and the CAT model performance does increase.

6.5.2 Assessing performance of URMOD and CAT on 15 Rodbourne events

This section will explore the difference in model performance based on modelling of the 15 events from the Rodbourne catchment listed in table 6.2. For this analysis both the NSE and MAE will be applied. In Figure 6-11 the left hand figure will be the difference in the NSE, and the right hand figure the difference in the MAE. The binomial hypothesis test will again be performed as a two-tailed test, testing both successes of either URMOD outperforms CAT or vice versa.

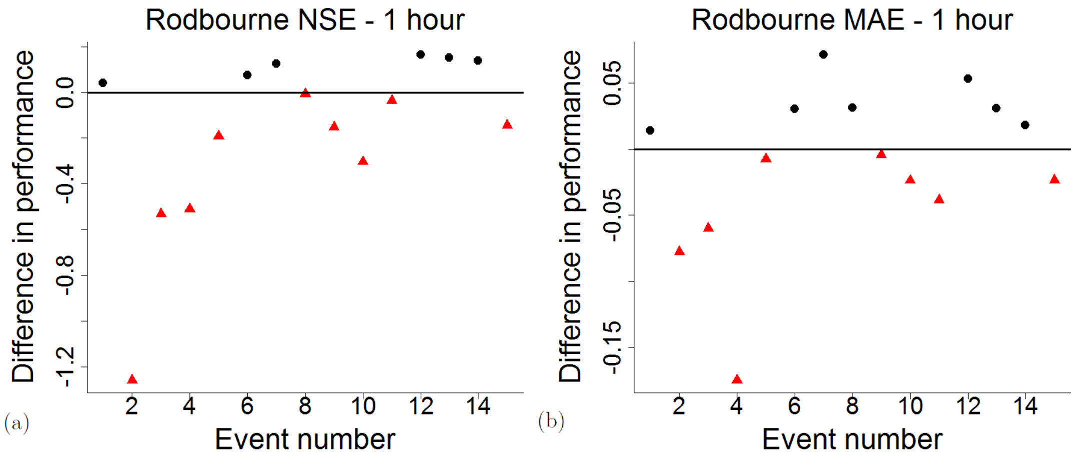


Figure 6-11: Difference in performance for the 15 events on the Rodbourne catchment on the hourly time step. The left figure (a) shows the NSE performance criteria and the right figure (b) MAE performance criteria. The black circles show that the CAT model performance is better than URMOD, whilst the red triangles show that performance of URMOD is better than CAT.

Figure 6-11 shows that the CAT model outperformed the URMOD model for six events out of a total of 15 when the NSE performance criteria is applied. Whilst the CAT model outperformed the URMOD model for seven events out of 15 when the MAE performance criteria is used. The MAE performance criteria differences indicate that 14 out of the 15 events are within 0.1, indicating similar performance between the models. Applying the binomial hypothesis test for both the NSE and MAE results in p-values of 0.61 and 1, respectively, indicating that the H_0 can be accepted in both instances, i.e no significant difference in model performance can be detached. Whilst the binomial approach and performance criteria explore performance for the entire event, the difference in observed and simulated peak flow performance will now be explored. Table 6.5 displays the observed peak flow, CAT and URMOD simulated peak and the difference between observed and simulated for both models.

Performance of peak flow of the URMOD and CAT model on the 15 Rodbourne events

Event number	Observed peak (m^3s^{-1})	CAT peak (m^3s^{-1})	Difference %	URMOD peak (m^3s^{-1})	Difference %
1	0.89	0.96	7.41	0.73	-22.29
2	0.64	0.27	-139.75	0.48	-33.86
3	0.99	0.88	-12.17	0.61	-61.94
4	1.43	0.89	-60.08	0.93	-53.49
5	0.71	0.78	9.06	0.72	2.28
6	1.08	1.06	-1.72	0.83	-30.30
7	1.13	1.15	1.82	0.96	-18.15
8	0.97	1.48	34.32	0.87	-11.36
9	0.84	1.18	28.2	0.74	-14.53
10	0.74	1.04	29	0.72	-2.64
11	1.06	0.82	-30.22	0.65	-64.83
12	0.46	0.87	46.6	0.8	41.94
13	0.72	0.68	-5.76	0.6	-18.94
14	0.92	1.02	10.56	0.83	-10.73
15	0.97	1.20	19.27	0.72	-35.38

Table 6.5: Observed, simulated and percentage difference in observed and simulated peak flow at each of the 15 events, using the URMOD and CAT models. Negative percentage difference shows a smaller peak by the model.

Table 6.5 shows that URMOD under-estimates the peak for 12 events and this is reduced to 6 events for the CAT model. One issue with the peak value simulated by the URMOD model is that they are not larger than $1m^3/s$. However the peak values for URMOD reported in Table 6.5 are the median value of fourteen simulations due to the jackknife calibration method. Both models difference in peak ranges are large with URMOD having a range value of 0.86 whilst the CAT model is 1.05. Whilst for the URMOD model the range is large due to a singular event (Event 12) with a difference of 0.46, once this is removed the range is 0.5. However the CAT does not have a singular value influencing the large range indicating that the CAT model is inconsistent.

Hydrograph representation of two events

The parameter set used for Figure 6-12 and 6-13 was event 1, which was classified as a short event. Figure 6-13 also shows a short event (Event 7), whereas Figure 6-12 displays an extended event (Event 2).

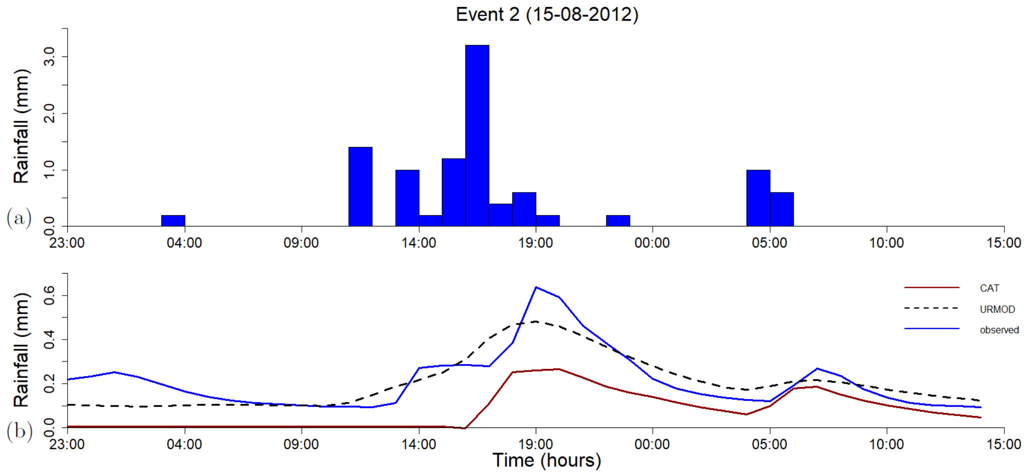


Figure 6-12: Event 2 hourly data, top figure (a) observed rainfall (mm), bottom figure (b) observed river flow (blue solid line), CAT estimated river flow (red solid line) and URMOD estimated river flow (black dashed line).

Figure 6-12 and 6-13 show observed and simulated flow from both the URMOD and CAT models. Figure 6-12 displays a relatively prolonged rainfall event, whereas Figure 6-13 contains a singular quick burst of rainfall. The results presented in Figure 6-13 show that both URMOD and CAT can model quicker rainfall events with both models matching the observed data, with URMOD under-estimating the peak. In contrast, the results presented in Figure 6-12 show that URMOD is capable of replicating the observed data with slight under estimation, whereas CAT is unable to represent the observed data with consistent under estimation.

URMOD parameter calibration results

Table 6.6 summarises the range of URMOD model parameter obtained from calibration across the 15 events on the Rodbourne catchment. The results show

	S	F	R	k	Φ	B_L	S_L	U_L	γ
Maximum	8044	1	1	1	42	289	3.16	3.16	0.25
Average	1574	0.67	0.4	0.61	11	97	2.67	2.43	0.15
Minimum	134	0.18	0.001	0.1	3	3	2.01	0.57	0

Table 6.6: URMOD parameter variability for 15, hourly Rodbourne events.

that the parameter range for URMOD is very large for all parameters except for S_L and γ. Interestingly the γ parameter has a small range alongside calibrating to a small value, indicating that the urban model wants to revert back to the rural model in some cases. These parameter ranges are within what has been generated from URMOD in previous chapters and studies, the S parameter is very large on average which was not expected since this value normally ranges between 100-1000, so having very large values was surprising. Similarly the B_L parameter was much larger than expected with values over 97 indicating that for some of the events the baseflow is time delayed longer than the entire event and so is not needed. One clear problem from this parameter set is generalisation between rainfall events would be not possible due to the very large range of values generated between each parameter. Since only a single initial condition set is used because differentiation between type of event is not considered then the parameter sets will have to account for the different climates (winter/ summer).

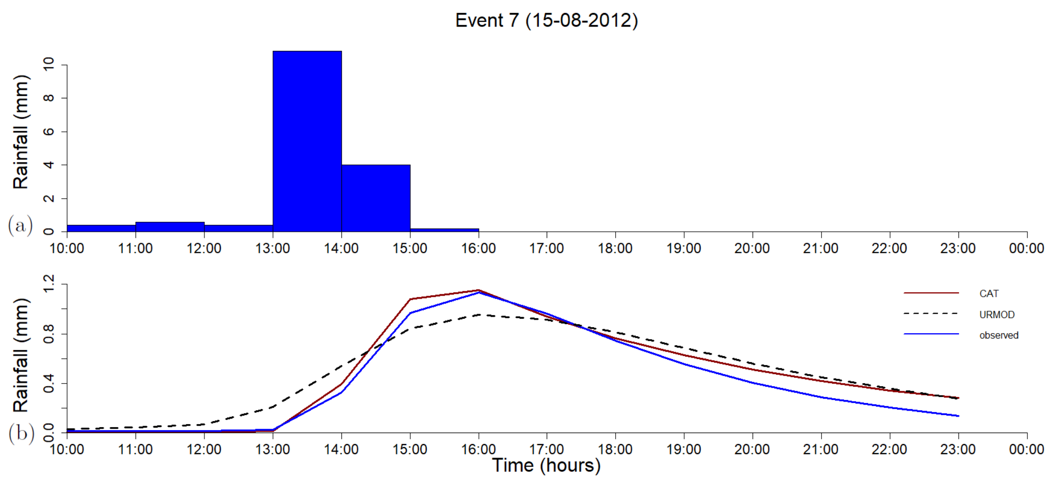


Figure 6-13: Event 7 hourly data, top figure (a) observed rainfall (mm), bottom figure (b) observed river flow (blue solid line), CAT estimated river flow (red solid line) and URMOD estimated river flow (black dashed line).

6.6 Discussion

The results presented in this paper were used to address the research question of whether a lumped urban rainfall-runoff model can simulate continuous and event based sub-daily data as accurately as a semi-distributed model. Results indicate that the lumped URMOD model is capable of outperforming the semi-distributed CAT model for the Gyeongan-Cheon catchment for 12-hourly and 3-hourly time steps, while no significant difference in performance could be detached for hourly time steps. The Rodbourne, event-based analysis showed that the two models performed equally well but URMOD did under-estimate on a number of peaks.

A review of the literature showed that when applying daily data the semi-distributed and distributed models should outperform the lumped models. This conclusion was shared in this paper as shown in Section 6.5.1. When daily data was applied the semi-distributed model CAT outperformed URMOD when comparing the NSE performance criteria but the reverse was not true when the MAE was applied. The main conclusion drawn from Figure 6-6 is that while model performance was similar the CAT model would be preferred due to URMOD under-estimating peaks.

The literature indicates that conflicting conclusions have been reached when comparing lumped and semi-distributed urban rainfall-runoff models using sub-daily data. The purpose of this paper was to further explore this problem, and did this through a continuous and event based analysis. The continuous analysis conducted on the Gyeongan-Cheon catchment compared model performance when a large volume of rainfall was present. Conflicting conclusions, dependent on the time step, used were found. For the 12-hourly and 3-hourly data it is shown using the Binomial hypothesis test URMOD is statistically significantly better than CAT. However when hourly datas applied, the binomial hypothesis test showed that model performance were not statistically significantly different. This conclusion on the hourly data is in line with the findings of Dotto et al. (2011). When exploring simulated output from both models its shown that for 12-hourly and 3-hourly data the CAT model consistently over and under-estimate observed streamflow. Whilst URMOD was capable of matching the observed flow with minimal under-estimation expect during extremely large rainfall events. Neither model performed well when hourly data was applied. One key point to make is that there was missing data for the sub-daily data for Gyeongan-Cheon catchment. This has a larger effect on URMOD than the CAT model due to URMOD having to calibrate parameters. However as shown in the difference of performance figures, Binomial hypohtesis test and the hydrographs URMOD does still outperform the CAT model for every time step.

It was expected that the CAT model to outperform URMOD on the Gyeongan-Cheon catchment for both daily and sub-daily time steps. Firstly due to the CAT model being a more detailed model than URMOD but as indicated in the literature conflicting conclusions do exist. Secondly URMOD was not initially designed to be applied to catchments with such high seasonality like the Gyeongan-Cheon catchment. As expected the CAT model did outperform URMOD when the daily data is applied, however questions of performance are raised when the sub-daily data is used. Whilst the NSE shows that URMOD did out-perform the CAT model when the sub-daily data's applied, the hydrographs showed that neither model was capable of matching the peak flows. For a practical application the CAT model would be preferred when daily data is used but for sub-daily further research will be needed in order to better parameterise URMOD.

The event based analysis was conducted on the Rodbourne catchment, comparing both short and extended rainfall events. When applying performance criteria its shown that the URMOD model was preferred, however the Binomial hypothesis test showed no statistical significance. One important issue that needs to be addressed is model calibration. The physically based CAT model only had a single set of parameters applied to it. Whilst URMOD was calibrated on each event and a median performance criteria was obtained for every event. This could have influenced the results due to two different types of events chosen (short and extended) but it did not, indicating that URMOD is potentially a lot more flexible than CAT in being able to simulate for both types of events. However, as shown in Table 6.6 URMOD's parameter variability is extremely large, which means generalisation of URMOD parameters being impossible, unlike CAT. However, due to a single parameter set being selected for CAT, this meant that it was unable to estimate both prolonged and flashy rainfall events, as shown in Figure 6-12 and 6-13.

6.7 Conclusion

This study uses a combination of hydrographs, performance criteria and binomial hypothesis tests in order to compare a lumped conceptual rainfall-runoff model and a semi distributed physically based hydrological model. The results showed that a simpler lumped conceptual model is preferable to a physical semi-distributed model, when computational time is not a factor. The CAT models computational time is a few seconds since no calibration is needed, unlike URMOD which required approximately 20-30 minutes calibration time for the Gyeongan-Cheon catchment and 5-10 minutes for the Rodbourne catchment. The ability to generate multiple parameter sets with ease means that URMOD is more flexible than the CAT model. Currently work is being done at KICT to create a calibration method for the CAT model which will give it the ability to have more accurate parameters. But there is currently no way to implement multiple parameter sets into one run of the CAT model unless it is done over multiple runs with stopping and physically changing the parameters. However a slight trade off in performance needs to be made since the URMOD model does underestimate peak events. If more detailed land use data is available and flow within a catchment is needed then CAT is a better model due to URMOD not being able to generate these results. This conclusion conflicts with the statement made by Salvadore et al. (2015) that distributed models are the future of urban modelling. As such future research needs to be conducted on the ability of lumped urban rainfall-runoff models to simulate using sub-daily data.

Acknowledgments

The NERC funded POLLCURB project for providing access to the hydrological and land-use data used in this study (NE/K002317/1). NERC for access to land-use shapefiles, Data owned by NERC Centre for Ecology & Hydrology. The Rodbourne figure contains public sector information licensed under the Open Government Licence v3.0.

6.A CAT and URMOD parameter tables

Parameter	Description	Unit
Area	Area	km^2
Slope	Slope	%
Aratio_imp	Impervious area ratio	%
Aratio_per	Pervious area ratio	%
Aratio_per_plant	Vegetation area ratio in pervious	%
depC_imp	Depression capacity in impervious zone	mm
depC_per	Depression capacity in pervious zone	mm
θ_per	Current soil moisture content	-
Soil_th_per	Soil thickness	m
s_per	Saturated soil moisture content	-
r_per	Residual soil moisture content	-
FC_per	Soil moisture content at field capacity	-
W_per	Soil moisture content at wilting point	-
ks_per	Saturated hydraulic conductivity	mms^{-1}
ksi_per	Saturated horizontal hydraulic conductivity	mms^{-1}
Mualem	Index for Mualem equation	-
gwE	Current groundwater level	m
rivE	Riverbed elevation	m
riv_th	Thickness of riverbed	m
ku_riv	Saturated hydraulic conductivity of riverbed	mms^{-1}
Area_riv	Area of riverbed	km^2
aqf_S	Storage coefficient of aquifer	-
aqf_Top	Top elevation of aquifer	m
aqf_Bot	Bottom elevation of aquifer	m
GW pump_rate	Groundwater pumping rate	day
leakage_rate	Leakage rate of water supply network	-
connect	Groundwater movement connecting node	-
aquifer slope	Average slope of groundwater in aquifer	-
node length	Average node length	-
conj. Length	Contact zone length between contiguity watershed	m
Kgw	Saturated hydraulic conductivity of aquifer	mms^{-1}

Table 6.7: CAT model parameters, description and units

Parameter	Description	Unit
S	Average Soil moisture capacity	mm
F	Field capacity	mm
R	Rooting Depth	mm
k	Drainage coefficient	-
Φ	Proportion of runoff split	-
B_L	Baseflow lag	day
S_L	Channel lag	day
U_L	Urban lag	day
γ	Scaling term	-

Table 6.8: URMOD model parameters, description and units

6.2 Bibliography

Ajami, N. K., Gupta, H., Wagener, T., and Sorooshian, S. (2004). Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system. *Journal of Hydrology*, 298(1-4):112–135.

Atkinson, S., Sivapalan, M., Woods, R., and Viney, N. (2003). Dominant physical controls on hourly flow predictions and the role of spatial variability: Mahurangi catchment, New Zealand. *Advances in Water Resources*, 26(3):219–235.

Bayliss, A., Black, K., Fava-Verde, A., and Kjeldsen, T. (2006). URBEXT2000 - a new FEH catchment descriptor. calculation, dissemination and application. Technical Report R&D FD1919/TR, Department for Environment Food and Rural Affairs, CEH wallingford.

Beven, K. J. (2011). *Rainfall-runoff modelling: the primer*. John Wiley & Sons, Chichester.

Chow, V.T, M. D. and Mays, L. (2010). *Applied Hydrology*. McGraw-Hill series in water resources and environmental engineering. Tata McGraw-Hill Education.

Cunge, J. A. (1969). On the subject of a flood propagation computation method (musklngum method). *Journal of Hydraulic Research*, 7(2):205–230.

Das, T., Bárdossy, A., Zehe, E., and He, Y. (2008). Comparison of conceptual model performance using different representations of spatial variability. *Journal of Hydrology*, 356(1-2):106–118.

Dotto, C. B., Kleidorfer, M., Deletic, A., Rauch, W., McCarthy, D. T., and Fletcher, T. D. (2011). Performance and sensitivity analysis of stormwater models using a bayesian approach and long-term high resolution data. *Environmental Modelling & Software*, 26(10):1225–1239.

Duan, Q., Gupta, V. K., and Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Applications*, 76(3):501–521.

El-Nasr, A. A., Arnold, J. G., Feyen, J., and Berlamont, J. (2005). Modelling the hydrology of a catchment using a distributed and a semi-distributed model. *Hydrological Processes*, 19(3):573–587.

Green, W. H. and Ampt, G. (1911). Studies on soil phyics. part I. the flow of air and water through soils,. *The Journal of Agricultural Science*, 4(1):1–24.

Horton, R. E. (1933). The role of infiltration in the hydrologic cycle. *Eos, Transactions American Geophysical Union*, 14(1):446–460.

Jang, C. H., Kim, H. J., Ahn, S. R., and Kim, S. J. (2016). Assessment of hydrological changes in a river basin as affected by climate change and water management practices, by using the CAT model. *Irrigation and Drainage*, 65:26–35.

Jia, Y. (1997). *Integrated analysis of water and heat balances in Tokyo metropolis with a distributed model.* PhD thesis, University of Tokyo, Japan.

Kim, H.-J., Jang, C.-H., and Noh, S.-J. (2012). Development and application of the catchment hydrologic cycle assessment tool considering urbanization (I)-model development. *Journal of Korea Water Resources Association*, 45(2):203–215.

Klemeš, V. (1986). Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, 31(1):13–24.

Krysanova, V., Bronstert, A., and Müller-Wohlfeil, D.-I. (1999). Modelling river discharge for large drainage basins: from lumped to distributed approach. *Hydrological Sciences Journal*, 44(2):313–331.

Li, H., Xu, C.-Y., and Beldring, S. (2015). How much can we gain with increasing model complexity with the same model concepts? *Journal of Hydrology*, 527:858–871.

Lighthill, M. J. and Whitham, G. (1955). On kinematic waves i. flood movement in long rivers. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 229(1178):281–316.

Linsley, R. K., Kohler, M. A., and Paulhus, J. L. (1949). *Applied hydrology.* The McGraw-Hill Book Company, Inc.; New York.

Miller, J. D., Kim, H., Kjeldsen, T. R., Packman, J., Grebby, S., and Dearden, R. (2014). Assessing the impact of urbanization on storm runoff in a peri-urban catchment using historical change in impervious cover. *Journal of Hydrology*, 515:59–70.

Monteith, J. L. (1965). Evaporation and environment, in the state and movement of water in living organisms. In *Symp. Soc. Exp. Biol.*, pages 205–234. Academic Press.

Nash, J. E. and Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I-A discussion of principles. *Journal of Hydrology*, 10(3):282–290.

Rawls, W. J. and Brakensiek, D. (1985). Prediction of soil water properties for hydrologic modeling. In *Watershed management in the eighties*, pages 293–299. ASCE.

Refsgaard, J. C. and Knudsen, J. (1996). Operational validation and intercomparison of different types of hydrological models. *Water Resources Research*, 32(7):2189–2202.

Salvadore, E., Bronders, J., and Batelaan, O. (2015). Hydrological modelling of urbanized catchments: a review and future directions. *Journal of Hydrology*, 529:62–81.

Ward, H., Evans, J. G., and Grimmond, C. S. B. (2013). Multi-season eddy covariance observations of energy, water and carbon fluxes over a suburban area in swindon, uk. *Atmospheric Chemistry and Physics*, 13(9):4645–4666.

Chapter 7

Conclusion

The purpose of this thesis is to create a widely applicable and simple-to-use framework for modelling the impact of urbanisation on catchment runoff using lumped conceptual rainfall-runoff models. This can be seen as an effect to reduce model structure error (reduce ϵ_2) when modelling urban catchments. In order to answer the overall aim three research question were formed, each of them testing a different part of the overall aim of the thesis. The results will be discussed in detail with reference to each research questions and an overall conclusion will be drawn. The first research question to be explored is research question 3;

Research question 3: Can simple statistical-based techniques be used to improve validation of rainfall-runoff models.

This hypothesis was explored in Chapter 4: Operational model comparison techniques for rainfall-runoff models. The purpose of this research question was to develop robust model performance techniques in order to reduce the impact of model calibration error (ϵ_3). Two model comparison techniques are developed, these were a paired z-test and a binomial hypothesis test. Alongside these two model comparison tools a jackknife model calibration and validation method was designed. The jackknife model calibration and validation method was applied in Chapters 4, 5 and 6 and shown to have many advantages and only one disadvantage when compared to traditional split-sample tests.

The first advantage is that the method is flexible, it can be applied to yearly or monthly periods to calibrate the model or as shown in Chapter 6 an event based calibration. Secondly, since the method uses the entire data set inconsistencies in the data can be found, such that if a certain period is consistently under-performing through multiple jackknife calibrations this under-performance can be addressed. This is opposed to a simple split-sample test, which could not have detected it. Alongside this, due to a larger number of model calibrations relative to a singular split-sample test, more sets of model parameters can be generated.

Since more parameter sets are generated, initial model parameter analysis can be undertaken. This would be achieved by exploring the range of the parameters to determine if one or multiple have the same value across multiple calibrations and catchments. Whilst more advanced tools do exist to explore parameter significance this method is not supposed to replace sensitivity analysis but be an initial sensitivity analysis before the more advanced tools are used. The main advantage of the method is that it utilises the entire data and can generate multiple performance criteria which can then be used for further analysis. However this creates the disadvantage of the jackknife hypothesis testing methodology such there is an increase in computational time (standard computational time is 30 minutes for a single 30 year split-sample calibration with 10000 iterations). Depending on the number of jackknife iterations the increase in computational time can be up to 20x longer than a split-sample test. The paired z-test results in Chapter 4 showed that for all catchments nether model was statistically significantly better than the other. Further testing and analysis is needed in order to determine if the paired z-test can be applied and due to it being outside of the scope of this thesis was not further explored.

Unlike the z-test the binomial hypothesis test tool developed in Chapter 4 was shown to be successful due to it creating a methodology to compare multiple catchments or events. The main advantage of the binomial hypothesis test is that it can assess comparative model performance across a large number of catchments or, as shown in Chapter 6, events. This indicates that the method is flexible as it can be applied to a number of different situations. The binomial test is also flexible when used in conjunction with the jackknife calibration and validation method as shown in Chapter 5, by using the multiple validation periods generated from the output of the jackknife method. One disadvantage of the binomial test is a minimum number of data points is needed for it to be applied, but can be accounted for when planning on applying the test. Research question 1 applied the binomial hypothesis test and jackknife calibration and validation method developed in order to answer the question;

Research question 1: Can rainfall-runoff modelling in urban catchments be improved via implementation of conceptual urban framework onto an already existing rural based model?

Three urban runoff generation frameworks and two routing frameworks were designed and tested in order to develop a conceptual urban framework that could reduce the model structure error (ϵ_2) in urban rainfall-runoff modelling. Following initial investigation, urban framework 3 was selected. Two routing methodologies were also developed, one which routed through a linear reservoir and the second a bounded pipe network. The urban routing methodology, alongside urban framework 3, were combined to make URMOD. A comparison was undertaken between URMOD and the default rural model using daily data on 27 catchments located within the Thames catchment. Results showed that for a 1 or 2-year calibration period, when applying the paired z-test performance was very similar for individual catchment comparisons indicating that potentially the urban framework did not add enough explanatory power to the model description of the catchment hydrology. However when the binomial hypothesis test was applied it showed that the two models collective performance of the catchments differed depending on the calibration time step used. For yearly calibration significant difference in model performance was seen, however for 2-yearly results no difference in performance was seen. As stated in the literature review a number of studies concluded that urban flow has traveled through the catchment system within a day. This indicates that the urban routing system in the model was potentially hindering model performance and not improving it.

As such a more detailed analysis of two catchments with sub-daily data was presented in Chapter 5 and 6. The results in Chapter 5 showed that for the catchment of the river Ray (NRFA 39087) URMOD outperformed the rural model in every time step, except the 8 hour time step when performance criteria and the Binomial hypothesis test was applied. When exploring hydrograph performance of the two models, URMOD was shown to outperform the rural model. The results showed that the rural model was unable to match the baseflow, showing consistent under-estimation for all time steps. In contrast, URMOD was able to capture the baseflow and only slightly over-estimate peak flows. This could be a result of the urban routing in the model meaning the baseflow aspect of URMOD had to account for less of the total flow than DAYMOD.

Questions of performance were raised when comparing the two models for the River Cut (NRFA 39052), this is due to the rural model outperforming URMOD for each time step when applying the NSE criteria, but only two time steps when applying the MAE criteria. However exploring hydrograph performance showed that the rural model consistently under-simulated flow, whereas URMOD was capable of matching peak flows and rainfall events. However neither model was capable of matching the baseflow during the 1990-1999 period. The conclusion that was drawn was that whilst the rural model performed well the URMOD model performance was just as good. Overall the URMOD model is a better performing model than the default rural model on both daily and sub-daily data when applied to urban catchments. Whilst some questions of performance are raised due to URMOD not outperforming the rural model in every occasion, the rural model is a nested version of URMOD and due to the human influence of urban catchments at times the urban framework may potentially be ineffective. This is because the urban extension may be trying to account for a urban signal which is not present. Chapter 6 presented an comparison between URMOD and the CAT model in order to test the performance of a lumped model against a semi-distributed model.

Research question 2: What is the loss of information when moving from lumped to spatially distributed rainfall-runoff models on a varying time scale for urban catchments?

This hypothesis was explored in Chapter 6, by comparing a lumped model (URMOD) and a semi-distributed model (CAT). The analysis was conducted on two catchments, Rodbourne and the Gyeongan-cheon both catchments were selected for specific reasons. Rodbourne was selected due to it being a smaller catchment (5.5 km^2) with a large percentage of urbanisation (46%). The Gyeongan-cheon catchment was selected due to it being a larger catchment (260.1 km^2) and having very high seasonal rainfall. The reason for these two catchments was to test the capabilities of the URMOD model alongside the CAT model. Similarly the comparison of continuous simulations and event based analysis on sub-daily data was again chosen to test the performance of URMOD. Results showed that for the continuous simulations URMOD was a better performing model, outperforming CAT for 12-hourly and 3-hourly data when applying the binomial hypothesis test. However hourly simulations was reported as similar performance. However when exploring the hydrograph performance, results showed that for hourly data the URMOD model was unable to match a large rainfall event. The event based analysis showed that when the Binomial hypothesis test was applied neither model out-performed the other. The URMOD models average performance was very good across all events with all parameter sets generated, however issues were raised as URMOD did under estimate the peaks in a number of catchments which is a consistent problem for the model. This was a problem that was not shared by the CAT model. Another issue raised by the event based analysis was that the URMOD models parameter range was very large indicating that generalisation of parameters would be incredibly difficult if not impossible. Overall results showed that URMOD was a better performing that the CAT model, but improvements to the URMOD model can still be made. Improvements such as extra model processes can be added due to the building block style of the model.

All three of the research questions were used in order to answer the main aim of the thesis. The overall conclusion of this work is that URMOD is a good start towards a better simple methodology of capturing urban effects on river flow in urban catchments. Whilst further work is still needed to improve the model, currently it has the capability to outperform an existing rural and urban model with a high level accuracy. The thesis has achieved the aim of attempting to reduce the error ϵ_2 in an urban model framework.

Chapter 8

Bibliography

Abbott, M., Bathurst, J., Cunge, J., O'connell, P., and Rasmussen, J. (1986). An introduction to the european hydrological system - systeme hydrologique europeen, "she", 2: Structure of a physically-based, distributed modelling system. *Journal of Hydrology*, 87(1-2):61–77.

Achleitner, S., Möderl, M., and Rauch, W. (2007). City drain©–an open source approach for simulation of integrated urban drainage systems. *Environmental Modelling & Software*, 22(8):1184–1195.

Ahn, K.-H. and Merwade, V. (2014). Quantifying the relative impact of climate and human activities on streamflow. *Journal of Hydrology*, 515:257–266.

Ajami, N. K., Gupta, H., Wagener, T., and Sorooshian, S. (2004). Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system. *Journal of Hydrology*, 298(1-4):112–135.

Andréassian, V., Perrin, C., Berthet, L., Le Moine, N., Lerat, J., Loumagne, C., Oudin, L., Mathevet, T., Ramos, M., and Valéry, A. (2009). Crash tests for a standardized evaluation of hydrological models. *Hydrology and Earth System Sciences Discussions*, (13):p–1757.

Anh, N. L., Boxall, J., Saul, A., and Willems, P. (2010). An evaluation of three lumped conceptual rainfall-runoff models at catchment scale. *Proceedings of the 3rd International Symposium on British Hydrological Society, Newcastle, UK, July.*

Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R. (1998). Large area hydrologic modeling and assessment part I: model development. *JAWRA Journal of the American Water Resources Association*, 34(1):73–89.

Askew, A. J. (1970). Derivation of formulae for variable lag time. *Journal of Hydrology*, 10(3):225–242.

Atkinson, S., Sivapalan, M., Woods, R., and Viney, N. (2003). Dominant physical controls on hourly flow predictions and the role of spatial variability: Mahurangi catchment, New Zealand. *Advances in Water Resources*, 26(3):219–235.

Basnayaka, A. and Sarukkalige, P. R. (2011). Comparing hydrology and hydraulics surface routing approaches in modeling an urban catchment. In *Proceedings of the 2nd International Conference on Environmental Engineering and Applications*, pages 123–127. IACSIT Press.

Bayliss, A., Black, K., Fava-Verde, A., and Kjeldsen, T. (2006). URBEXT2000 - a new FEH catchment descriptor. calculation, dissemination and application. Technical Report R&D FD1919/TR, Department for Environment Food and Rural Affairs, CEH wallingford.

Beck, M. B. (1991). Forecasting environmental change. *Journal of Forecasting*, 10(1-2):3–19.

Bergstrom, S. (1976). Development and application of a conceptual runoff model for scandinavian catchments. SMHI Norrköping, Report RH07.

Beven, K., Kirkby, M., Schofield, N., and Tagg, A. (1984). Testing a physically-based flood forecasting model (TOPMODEL) for three UK catchments. *Journal of Hydrology*, 69(1-4):119–143.

Beven, K. J. (2011). *Rainfall-runoff modelling: the primer*. John Wiley & Sons, Chichester.

Bhaskar, A. S., Welty, C., Maxwell, R. M., and Miller, A. J. (2015). Untangling the effects of urban development on subsurface storage in Baltimore. *Water Resources Research*, 51(2):1158–1181.

Bouffard, J.-S. (2014). *A comparison of conceptual rainfall-runoff modelling structures and approaches for hydrologic prediction in ungauged peatland basins of the James Bay lowlands*. PhD thesis, Carleton University Ottawa. Retrieved from https://curve.carleton.ca/7ed3b15f-bfff-4027-9a87-f314602d7a1a.

Boughton, W. and Chiew, F. (2007). Estimating runoff in ungauged catchments from rainfall, PET and the AWBM model. *Environmental Modelling & Software*, 22(4):476–487.

Boyd, M., Rigby, E., and Vandrie, R. (1996). WBNM a comprehensive flood model for natural and urban catchments. In *Proc., 7th Int. Conf. on Urban Drainage, Institution of Engineers*, pages 329–334.

Boyle, D. P., Gupta, H. V., Sorooshian, S., Koren, V., Zhang, Z., and Smith, M. (2001). Toward improved streamflow forecasts: Value of semidistributed modeling. *Water Resources Research*, 37(11):2749–2759.

Brun, S. and Band, L. (2000). Simulating runoff behavior in an urbanizing watershed. *Computers, Environment and Urban Systems*, 24(1):5–22.

Butts, M. B., Payne, J. T., Kristensen, M., and Madsen, H. (2004). An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *Journal of Hydrology*, 298(1):242–266.

Chiew, F. and McMahon, T. (1994). Application of the daily rainfall-runoff model MODHYDROLOG to 28 australian catchments. *Journal of Hydrology*, 153(1-4):383–416.

Clark, M. P., Kavetski, D., and Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research*, 47(9):301.

Crawford, N. H. and Linsley, R. K. (1966). Digital simulation in hydrology'stanford watershed model 4. Technical Report Technical Report,No. 39, Stanford University, Palo Alto, California.

Criss, R. E. and Winston, W. E. (2008). Do Nash values have value? Discussion and alternate proposals. *Hydrological Processes*, 22(14):2723.

Cunge, J. A. (1969). On the subject of a flood propagation computation method (musklngum method). *Journal of Hydraulic Research*, 7(2):205–230.

Das, T., Bárdossy, A., Zehe, E., and He, Y. (2008). Comparison of conceptual model performance using different representations of spatial variability. *Journal of Hydrology*, 356(1-2):106–118.

DeFries, R. and Eshleman, K. N. (2004). Land-use change and hydrologic processes: A major focus for the future. *Hydrological Processes*, 18(11):2183–2186.

Devia, G. K., Ganasri, B., and Dwarakish, G. (2015). A review on hydrological models. *Aquatic Procedia*, 4:1001–1007.

Diem, J. E., Hill, T. C., and Milligan, R. A. (2018). Diverse multi-decadal changes in streamflow within a rapidly urbanizing region. *Journal of Hydrology*, 556:61–71.

Dominion, F. (1999). Complexity and the commons. *Simon Levin. Reading, MA: Perseus Books*.

Donnelly-Makowecki, L. and Moore, R. (1999). Hierarchical testing of three rainfall–runoff models in small forested catchments. *Journal of Hydrology*, 219(3):136–152.

Dotto, C. B., Kleidorfer, M., Deletic, A., Rauch, W., McCarthy, D. T., and Fletcher, T. D. (2011). Performance and sensitivity analysis of stormwater models using a bayesian approach and long-term high resolution data. *Environmental Modelling & Software*, 26(10):1225–1239.

Duan, Q., Gupta, V. K., and Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Applications*, 76(3):501–521.

Džubáková, K. (2010). Rainfall-runoff modelling: Its development, classification and possible applications. *ACTA Geographica universitatis comenianae*, 54(2):173–181.

El-Nasr, A. A., Arnold, J. G., Feyen, J., and Berlamont, J. (2005). Modelling the hydrology of a catchment using a distributed and a semi-distributed model. *Hydrological Processes*, 19(3):573–587.

Ewen, J. (2011). Hydrograph matching method for measuring model performance. *Journal of Hydrology*, 408(1):178–187.

Ewen, J. and O'Donnell, G. (2012). Prediction intervals for rainfall–runoff models: raw error method and split-sample validation. *Hydrology Research*, 43(5):637–648.

Fletcher, T., Andrieu, H., and Hamel, P. (2013). Understanding, management and modelling of urban hydrology and its consequences for receiving waters: A state of the art. *Advances in Water Resources*, 51:261–279.

Franczyk, J. and Chang, H. (2009). The effects of climate change and urbanization on the runoff of the rock creek basin in the portland metropolitan area, oregon, usa. *Hydrological Processes*, 23(6):805–815.

Gharari, S., Hrachowitz, M., Fenicia, F., and Savenije, H. (2013). An approach to identify time consistent model parameters: sub-period calibration. *Hydrology and Earth System Sciences, 17, 2013*.

Goyen, A., Phillips, B., and Black, D. (1991). Recent advances in flood estimation using rafts-xp. In *Proceedings of the International Hydrology and Water Resources Symposium, vol. 1, Institution of Engineers, Australia, Canberra*, pages 66–71.

Grayson, R. B., Blöschl, G., Western, A. W., and McMahon, T. A. (2002). Advances in the use of observed spatial patterns of catchment hydrological response. *Advances in Water Resources*, 25(8-12):1313–1334.

Gregory, J. H., Dukes, M. D., Jones, P. H., and Miller, G. L. (2006). Effect of urban soil compaction on infiltration rate. *Journal of Soil and Water Conservation*, 61(3):117–124.

Guo, J. C.-Y. (2006). *Urban hydrology and hydraulic design*. Water Resources Publication.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F. (2009). Decomposition of the mean squared error and nse performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1-2):80–91.

Hawley, R. J. and Bledsoe, B. P. (2011). How do flow peaks and durations change in suburbanizing semi-arid watersheds? a southern California case study. *Journal of Hydrology*, 405(1-2):69–82.

Hooper, R. P. (2001). Applying the scientific method to small catchment studies: a review of the Panola Mountain experience. *Hydrological Processes*, 15(10):2039–2050.

Houston, D., Werrity, A., Bassett, D., Geddes, A., Hoolachan, A., and McMillan, M. (2011). Pluvial (rain-related) flooding in urban areas: the invisible hazard.

Hsieh, L.-S. and Wang, R.-Y. (1999). A semi-distributed parallel-type linear reservoir rainfall-runoff model and its application in Taiwan. *Hydrological Processes*, 13(8):1247–1268.

Hydraulics, W. D. (2005). Sobek river/estuary user manual.

Jacobson, C. R. (2011). Identification and quantification of the hydrological impacts of imperviousness in urban catchments: A review. *Journal of Environmental Management*, 92(6):1438–1448.

Jain, S. K. and Sudheer, K. (2008). Fitting of hydrologic models: a close look at the nash–sutcliffe index. *Journal of Hydrologic Engineering*, 13(10):981–986.

Jakeman, A. and Hornberger, G. (1993). How much complexity is warranted in a rainfall-runoff model? *Water resources research*, 29(8):2637–2649.

Jones, J. A., Swanson, F. J., Wemple, B. C., and Snyder, K. U. (2000). Effects of roads on hydrology, geomorphology, and disturbance patches in stream networks. *Conservation Biology*, 14(1):76–85.

Kampf, S. K. and Burges, S. J. (2007). A framework for classifying and comparing distributed hillslope and catchment hydrologic models. *Water Resources Research*, 43(W05423).

Keller, V., Tanguy, M., Prosdocimi, I., Terry, J., Hitt, O., Cole, S., Fry, M., Morris, D., and Dixon, H. (2015). CEH-GEAR: 1 km resolution daily and monthly areal rainfall estimates for the UK for hydrological and other applications. *Earth System Science Data*, 7(1):143–155.

Kidd, C. and Lowing, M. (1979). The wallingford urban subcatchment model. Technical Report IH Report No.60 (unpublished), Institute of Hydrology, CEH Wallingford.

Kim, H.-J., Jang, C.-H., and Noh, S.-J. (2012). Development and application of the catchment hydrologic cycle assessment tool considering urbanization (I)-model development. *Journal of Korea Water Resources Association*, 45(2):203–215.

Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research*, 42(W03s04).

Kirchner, J. W., Hooper, R. P., Kendall, C., Neal, C., and Leavesley, G. (1996). Testing and validating environmental models. *Science of the Total Environment*, 183(1-2):33–47.

Kirkby, M. and Beven, K. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Journal*, 24(1):43–69.

Kjeldsen, T. (2009). Modelling the impact of urbanisation on flood runoff volumes. *Proceedings of ICE, Water Management*, 5:329–336.

Klemeš, V. (1986). Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, 31(1):13–24.

Krause, P., Boyle, D., and Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, 5:89–97.

Krysanova, V., Bronstert, A., and Müller-Wohlfeil, D.-I. (1999). Modelling river discharge for large drainage basins: from lumped to distributed approach. *Hydrological Sciences Journal*, 44(2):313–331.

Laurenson, E. M. (1964). A catchment storage model for runoff routing. *Journal of Hydrology*, 2(2):141–163.

Law, N. L., Cappiella, K., and Novotney, M. E. (2009). The need for improved pervious land cover characterization in urban watersheds. *Journal of Hydrologic Engineering*, 14(4):305–308.

Le Moine, N. (2008). *Le bassin versant de surface vu par le souterrain: une voie d'amélioration des performances et du réalisme des modèles pluie-débit?* PhD thesis, Université Pierre et Marie Curie.

Legates, D. R. and McCabe, G. J. (1999). Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, 35(1):233–241.

Li, H., Xu, C.-Y., and Beldring, S. (2015). How much can we gain with increasing model complexity with the same model concepts? *Journal of Hydrology*, 527:858–871.

Lighthill, M. J. and Whitham, G. (1955). On kinematic waves i. flood movement in long rivers. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 229(1178):281–316.

Linsley, R. K., Kohler, M. A., and Paulhus, J. L. (1949). *Applied hydrology.* The McGraw-Hill Book Company, Inc.; New York.

M, M. and J, W. (2004). *Modelling catchment hydrology.* Wiley.

Mays, L. W. and Tung, Y.-K. (2002). *Hydrosystems engineering and management.* Water Resources Publication.

McCuen, R. H., Knight, Z., and Cutter, A. G. (2006). Evaluation of the Nash–Sutcliffe Efficiency index. *Journal of Hydrologic Engineering*, 11(6):597–602.

McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of Mathematical Biophysics*, 7:115–133.

Miller, J. D., Kim, H., Kjeldsen, T. R., Packman, J., Grebby, S., and Dearden, R. (2014). Assessing the impact of urbanization on storm runoff in a peri-urban catchment using historical change in impervious cover. *Journal of Hydrology*, 515:59–70.

Mitchell, V., Mein, R. G., and McMahon, T. A. (2001). Modelling the urban water cycle. *Environmental Modelling & Software*, 16(7):615–629.

Moore, R. (1985). The probability-distributed principle and runoff production at point and basin scales. *Hydrological Sciences Journal*, 30(2):273–297.

Moore, R. and Bell, V. (2002). Incorporation of groundwater losses and well level data in rainfall-runoff models illustrated using the pdm. *Hydrology and Earth System Sciences Discussions*, 6(1):25–38.

Moradkhani, H. and Sorooshian, S. (2009). General review of rainfall-runoff modeling: model calibration, data assimilation, and uncertainty analysis. In *Hydrological Modelling and the Water Cycle*, pages 1–24. Springer.

Motiee, H., Chocat, B., and Blanpain, O. (1997). A storage model for the simulation of the hydraulic behaviour of drainage networks. *Water Science and Technology*, 36(8-9):57–63.

Mroczkowski, M., Raper, P. G., and Kuczera, G. (1997). The quest for more powerful validation of conceptual catchment models. *Water Resources Research*, 33(10):2325–2335.

Mulvaney, T. (1951). On the use of self-registering rain and flood gauges in making observations of the relations of rainfall and flood discharges in a given catchment. *Proceedings of the institution of Civil Engineers of Ireland*, 4(2):18–33.

Nash, J. E. and Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I-A discussion of principles. *Journal of Hydrology*, 10(3):282–290.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7(4):308–313.

Niemczynowicz, J. (1999). Urban hydrology and water management–present and future challenges. *Urban Water*, 1(1):1–14.

Ossola, A., Hahs, A. K., and Livesley, S. J. (2015). Habitat complexity influences fine scale hydrological processes and the incidence of stormwater runoff in managed urban ecosystems. *Journal of Environmental Management*, 159:1–10.

Packman, J. (1980). The effects of urbanisation on flood magnitude and frequency. Technical Report IH Report No. 63, Institute of Hydrology, CEH Wallingford.

Packman, J. (2004). Impact of antecedent rainfall longer than 5 days on base-flow and percentage runoff. Technical Report Environment agency R & D report W6-080, Centre for Ecology and Hydrology, CEH Wallingford.

Pechlivanidis, I., Jackson, B., McIntyre, N., and Wheater, H. (2011). Catchment scale hydrological modelling: a review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. *Global NEST Journal*, 13(3):193–214.

Pereira, C. (1982). Bernard augustus keen. 5 september 1890-5 august 1981. *Biographical Memoirs of Fellows of the Royal Society*, 28:205–223.

Perrin, C., Michel, C., and Andréassian, V. (2001). Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. *Journal of Hydrology*, 242(3):275–301.

Perrin, C., Michel, C., and Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, 279(1-4):275–289.

Price, G., Packman, J., and Kidd, C. (1980). A simplified model for sewered catchments. Technical Report IH Report No.62 (unpublished), Institute of Hydrology, CEH Wallingford.

Ragab, R., Rosier, P., Dixon, A., Bromley, J., and Cooper, J. (2003). Experimental study of water fluxes in a residential area: 2. road infiltration, runoff and evaporation. *Hydrological Processes*, 17(12):2423–2437.

Redfern, T. W., Macdonald, N., Kjeldsen, T. R., Miller, J. D., and Reynard, N. (2016). Current understanding of hydrological processes on common urban surfaces. *Progress in Physical Geography*, 40(5):699–713.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J., and Participants, D. (2004). Overall distributed model intercomparison project results. *Journal of Hydrology*, 298(1):27–60.

Refsgaard, J. C. (1997). Parameterisation, calibration and validation of distributed hydrological models. *Journal of Hydrology*, 198(1):69–97.

Refsgaard, J. C. and Henriksen, H. J. (2004). Modelling guidelines—-terminology and guiding principles. *Advances in Water Resources*, 27(1):71–82.

Refsgaard, J. C., Henriksen, H. J., Harrar, W. G., Scholten, H., and Kassahun, A. (2005). Quality assurance in model based water management–review of existing practice and outline of new approaches. *Environmental Modelling & Software*, 20(10):1201–1215.

Refsgaard, J. C. and Knudsen, J. (1996). Operational validation and intercomparison of different types of hydrological models. *Water Resources Research*, 32(7):2189–2202.

Rehman, H., Zollinger, M., Collings, G., et al. (2003). Hydrological vs hydraulic routing possibilities with two-dimensional hydraulic modelling. In *28th International Hydrology and Water Resources Symposium: About Water; Symposium Proceedings*, volume 3, pages 311–316. Institution of Engineers, Australia.

Ren-Jun, Z. (1992). The xinanjiang model applied in china. *Journal of hydrology*, 135(1-4):371–381.

Riad, S., Mania, J., Bouchaou, L., and Najjar, Y. (2004). Predicting catchment flow in a semi-arid region via an artificial neural network technique. *Hydrological Processes*, 18(13):2387–2393.

Rose, S. and Peters, N. E. (2001). Effects of urbanization on streamflow in the Atlanta area (Georgia, USA): a comparative hydrological approach. *Hydrological Processes*, 15(8):1441–1457.

Rosso, R. and Rulli, M. C. (2002). An integrated simulation method for flash-flood risk assessment: 2. effects of changes in land-use under a historical perspective. *Hydrology and Earth System Sciences*, 6(3):285–295.

Saint-Venant, A. d. (1871). Theorie du mouvement non permanent des eaux, avec application aux crues des rivieres et a l'introduction de marees dans leurs lits. *Comptes rendus des seances de l'Academie des Sciences*, 36:174–154.

Salvadore, E., Bronders, J., and Batelaan, O. (2015). Hydrological modelling of urbanized catchments: a review and future directions. *Journal of Hydrology*, 529:62–81.

Sargent, R. G. (2013). Verification and validation of simulation models. *Journal of simulation*, 7(1):12–24.

Schaefli, B. and Gupta, H. V. (2007). Do Nash values have value? *Hydrological Processes*, 21(15):2075–2080.

Seibert, J. (2003). Reliability of model predictions outside calibration conditions: Paper presented at the Nordic Hydrological Conference (Røros, Norway 4-7 August 2002). *Hydrology Research*, 34(5):477–492.

Shaw, E. (1994). *Hydrology in practice.* (3rd ed.), Chapman and Hall, London.

Sherman, L. K. (1932). Streamflow from rainfall by the unit-graph method. *Eng. News Record*, 108:501–505.

Shields, C. A. and Tague, C. L. (2012). Assessing the role of parameter and input uncertainty in ecohydrologic modeling: implications for a semi-arid and urbanizing coastal California catchment. *Ecosystems*, 15(5):775–791.

Shoemaker, L. (1997). *Compendium of tools for watershed assessment and TMDL development.* U.S. EPA, 1997.

Shuster, W., Bonta, J., Thurston, H., Warnemuende, E., and Smith, D. (2005). Impacts of impervious surface on watershed hydrology: a review. *Urban Water Journal*, 2(4):263–275.

Singh, V. P. et al. (1995). *Computer models of watershed hydrology.* Water Resources Publications Highlands Ranch, CO.

Sivapalan, M. and Woods, R. A. (1995). Evaluation of the effects of general circulation models' subgrid variability and patchiness of rainfall and soil moisture on land surface water balance fluxes. *Hydrological Processes*, 9(5-6):697–717.

Smith, J. and Eli, R. N. (1995). Neural-network models of rainfall-runoff process. *Journal of Water Resources Planning and Management*, 121(6):499–508.

Smith, M. B., Seo, D.-J., Koren, V. I., Reed, S. M., Zhang, Z., Duan, Q., Moreda, F., and Cong, S. (2004). The distributed model intercomparison project (DMIP): motivation and experiment design. *Journal of Hydrology*, 298(1):4–26.

Stephenson, D. (1994). Comparison of the water balance for an undeveloped and a suburban catchment. *Hydrological Sciences Journal*, 39(4):295–307.

Sudheer, K., Gosain, A., Mohana Rangan, D., and Saheb, S. (2002). Modelling evaporation using an artificial neural network algorithm. *Hydrological Processes*, 16(16):3189–3202.

Todini, E. (1996). The arno rainfall—runoff model. *Journal of hydrology*, 175(1-4):339–382.

Uhlenbrook, S., Roser, S., and Tilch, N. (2004). Hydrological process representation at the meso-scale: the potential of a distributed, conceptual catchment model. *Journal of Hydrology*, 291(3-4):278–296.

Valeo, C. and Moin, S. (2000). Grid-resolution effects on a model for integrating urban and rural areas. *Hydrological Processes*, 14(14):2505–2525.

Verbeiren, B., Van de Voorde, T., Canters, F., Binard, M., Cornet, Y., and Batelaan, O. (2013). Assessing urbanisation effects on rainfall-runoff using a remote sensing supported modelling strategy. *International Journal of Applied Earth Observation and Geoinformation*, 21:92–102.

Vogel, R. M. and Sankarasubramanian, A. (2003). Validation of a watershed model without calibration. *Water Resources Research*, 39(10).

Wagener, T., Wheater, H., and Gupta, H. V. (2004). *Rainfall-runoff modelling in gauged and ungauged catchments*. World Scientific.

Wang, Z.-M., Batelaan, O., and De Smedt, F. (1996). A distributed model for water and energy transfer between soil, plants and atmosphere (wetspa). *Physics and Chemistry of the Earth*, 21(3):189–193.

Weglarczyk, S. (1998). The interdependence and applicability of some statistical quality measures for hydrological models. *Journal of Hydrology*, 206(1-2):98–103.

WHO (2014). World Health Organisation urban population growth. `http://www.who.int/gho/urban_health/situation_trends/urban_population_growth_text/en/`. Accessed: 2016-03-25.

Wiles, T. J. and Sharp, J. M. (2008). The secondary permeability of impervious cover. *Environmental & Engineering Geoscience*, 14(4):251–265.

Willmott, C. J. and Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1):79–82.

Wright, N. G., Dastorani, M. T., Goodwin, P., and Slaughter, C. W. (2002). A combination of neural networks and hydrodynamic models for river flow prediction. In *Fifth International Conference on Hydroinformatics, Cardiff, U.K.*

Yang, D., Herath, S., and Musiake, K. (2000). Comparison of different distributed hydrological models for characterization of catchment spatial variability. *Hydrological Processes*, 14(3):403–416.

Yang, G., Bowling, L. C., Cherkauer, K. A., Pijanowski, B. C., and Niyogi, D. (2010). Hydroclimatic response of watersheds to urban intensity: an observational and modeling-based analysis for the white river basin, indiana. *Journal of Hydrometeorology*, 11(1):122–138.

Zhou, Y., Wang, Y., Gold, A. J., and August, P. V. (2010). Modeling watershed rainfall–runoff relations using impervious surface-area data with high spatial resolution. *Hydrogeology Journal*, 18(6):1413–1423.